# COMPUTATIONAL LINGUISTICS

# IV

COMPUTING CENTRE OF THE HUNGARIAN ACADEMY OF SCIENCES
BUDAPEST, 1965

# COMPUTATIONAL LINGUISTICS
# IV.

COMPUTING CENTRE OF THE HUNGARIAN ACADEMY OF SCIENCES
BUDAPEST, 1965

Publisher: The Computing Centre of the Hungarian Academy of Sciences

Addres: Budapest I., Uri utca 49

Felelős kiadó: FREY TAMÁS – Alak A/4 – Ívszám 23 A/5 ív
Megjelent 1965 novemberben 500 példányba
65300 M.T.A. KESz Sokszorosító üzem F. v.: Szabó Gyula

## C O N T E N T S


L. DEZSŐ page

Notes on the Word Order of Simple Sentences in Hungarian ... 3

S. ÁBRAHÁM

Some Questions of Phrase Structure Grammars I. .............61

F. KIEFER

Some Questions of Semantic Theory ........................ 71

Gy. SIPŐCZY

The Analysis of Prepositional Constructions ............... 79

E.N. FERENTZY

Computer Simulation of Human Behaviour and Concept Formation in Music Composition ................................. 93

E.N. FERENTZY

On Formal Music Analysis-Synthesis: Its Application in Music Education ..........................................107

K. CSÉBFALVI, M. HAVASS

A Direct Computer Processing of Folk Tunes ............... 125

APPENDIX

Reviews

С.К. Шаумян и П.А. Соболева:

Аппликативная порождающая модель и исчисление трансформаций в русском языке /S.J. Petőfi/ ...................... 133

F. Papp:

Matematikai nyelvészet és gépi forditás a Szovjetunióban /Mathematische Linguistik und Maschinenübersetzung in der Sowjetunion/ /S.J. Petőfi/ ............................ 163

R. Jakobson:

Selected Writings. Phonological Studies /Ф. Пап/ ....... 167

A. Szende:

Magyar nyelvtan és fogalmazás /Венгерская грамматика и сочинение/ - первый программированный учебник венгерского языке /Ф. Пап/ ......................................... 177

А. ЛЮДСКАНОВ

Машинный перевод в нвродной республике Болгарии ......... 183

# NOTES ON THE WORD ORDER OF SIMPLE SENTENCES IN HUNGARIAN [1]

L. Dezső

## 1. Preliminary notes

1.1 The study of Hungarian word order has a long history, rich in results, but for reasons of language it is scarecely known to international linguistic circles. Hungarian word order, which is free and yet subject to strict rules, has been a riddle not only to foreigners but also to Hungarian linguists for a long time. János Fogarasi was the first who discovered in his paper entitled "A szó helyezésről a magyar nyelvben" /published in 1838/ that the word preceding the verb /more precisely the verb without verbal prefix and the verb with separable verbal prefix/ is always stressed. Hústeszik Pali. A moziba Péter megy.[2] The investigations during the second half of the 19th century achieved further results, while following a partly different path. Following a paper starting the discussion of our greatest poet beside Petőfi, János Arany, some scholars including György Joannovics and Samuel Brassai contributed much to the exploration of this field. György Joannovics analyzed the word order properties of the different parts of speech and that of selected words like is, mindig, ritkán. Sámuel Brassai, a famous logician and linguist examined the word order problems of the known and the new in a message and his works in this field are of a pioneering character. The outstanding representative of the Hungarian "Junggrammatiker" Zsigmond Simonyi scrutinized the problems of word order on a vast historical material using also the results of his predecessors. The pertinent part of the grammar sponsored by the Hungarian Academy of Sciences was written by László Deme who enriched and systematized the results obtained so far. László Elekfi and Robert Hetzron have recently dealt with the problems of the logical segmentation of the sentence.

From the international literature on word order we want to mention abovr all the results obtained in the investigation of the word order of Slavic languages because the Slavic word order exhibits many common properties with those of Hungarian, and on the other hand word order plays just as important a role in these languages. E. Berneker scrutinized what places the different sentence-elements occupy in the sentence and his method of investigation was taken over by the authors of the most important handbooks. Following V. Mathesius the linguists of the Prague Circle: F. Travniček, V. Skalička, Fr. Daneš, P. Adamec and others[3] tackle various aspects of word order from the viewpoint of the actual segmentation. Among Soviet linguists K.G. Krušilneckaja and A.S. Melničuk follow the same path. In the field of the statistical investigation of the Russian word order the results of a German scholar, K. Buttke deserve attention. From the works on generative grammar we only want to mention the articles by E. Bach, D. Worth. In the near future another important work on German word order will be published by K. Heidolph whose work is a first attempt at the formal investigation of the actual segmentation within the compass of text theory. With respect to the typological investigation of word order special mention must be made of the works of J. Greenberg.

1.2 The present paper aims at examining some word order problems of simple sentences. But before starting the discussion we should delimit the scope of our investigations.

The analysis of word order can be carried out on two levels. On the first level only the order of words, their relations to each other are taken into consideration. If we consider the words as signs then this analysis is a synthetic one: we are interested in the rules concerning the placing of the sign in a linear sequence. Such an investigation can be made in two steps: analysis of the basic word order referring to the order of the nominal and verbal groups and within the verbal group to the order of the verb and its complements, the analysis of the internal word order of the nominal groups.

According to another word order analysis the problem is posed as to what may be regarded as new and what as already known concerning the message in the given speech situation. Such an analysis belongs to the field where language facts are investigated with respect to their dependence on the speech situation and can be called pragmatic analysis because it takes into consideration the conditions and circumstances of the uses of the signs. Since the speech situation can only be analyzed formally in a larger context than the sentence, such an investigation of word order comes under text analysis. Our paper tackles the Hungarian word order from a synthetic point of view but in 7. some problems concerning the interconnections between the syntactic and the pragmatic word order analysis are touched upon.

1.3. The next question to be answered is what we mean by simple sentence and what is its importance from the viewpoint of the word order analysis.

We do not use the term "simple sentence" in the traditional manner. We call such sentences simple as they contain nominal and predicative parts the nominal part consisting of a substantive and the predicative one - in the case of two elements - of an intransitive verb or a substantive without a copula /$N_n$/ or adjective /Adj/, in the case of three sentence elements the verb may be transitive or copula, the predicative part is complemented either by a substantive or an adverb. Only verbs without verbal prefixes may occur in simple sentences as understood here. Substantives may be preceded by an article except for the nominal predicate for which we do not permit the use of articles. If the substantives are complemented by an adjective /pronoun, numeral, $N_g$/, the order of these will be determined by special rules, namely by the word order rules of the nominal groups.

The sentences that may be produced in this way constitute only a small part of all possible sentences but a considerable part of the sentences with more than three elements may be generated from the elements of the sentences indicated above and other sentences with two and three elements may be obtained by means of pronominalization and ellipsis. The word order rules of the sentences with two or three but different elements may be derived from the word order rules established for the simple sentences if we apply some complementary rules.

The simple sentences are affirmative and declarative but interrogative and negative sentences will also be tackled briefly, though they are considered of secondary importance.

The problems of word order dealt with in this monograph constitute a part of the preliminary work towards a generative grammar of Hungarian. A survey of generative grammars does not come within the scope of the present paper. /The reader is referred to a recent book by Katz and Postal where an up-to-date bibliography of generative grammars can be found: Katz, J.J., Postal, P.M., An Integrated Theory of Linguistic Descriptions, M.I.T. Press, Cambridge, Mass. 1964/.

The arrangement of words can be carried out after the generation of the sentence only. If we alter a given sentence, i.e. if we change a declarative sentence to an interrogative sentence, an affirmative sentence to a negative one or if we alter the sentence to obtain an ellipsis then new word order rules may be valid.

1.4. The examination of word order rules has been carried out with the help of informants.[4] In the first phase of the analysis the informants had to determine whether a sentence written or uttered without stress accentuation is correct or not and if the answer was in the affirmative then the next question to be answered was which sentence element should be given logical stress. In the second phase of the analysis the same sentences were uttered in a way that an element was accented and the informants had to decide whether the sentence uttered in this way was correct or not. The two investigations were based on each other and one was taken as a control of the other to some extent.

Our task was to state the regularity in the answers of the informants. The statements obtained in this way were checked by means of text analysis. Summing up the result obtained with respect to the pattern sentences we tried to state the general rules of stressing and accentuation.

Before proceeding to analyze the sentence patterns and to establish the word order rules it seems worth while to clarify the relation between stress and accent. An element holds a stressed position in the sentence if it has logical prominence. Though the informants could decide this question, this criterion is not a formal one at all, but it is important from the viewpoint of the message so we have to make it more precise. As we will see below /in 5./, the stressed element receives the accent if there is only one accent in the sentence, so we exclude the cases of double accentuation with pauses. Such an accent does not require a complementing context after the sentence and this accent situation occurs in the independent sentences most frequently. In affirmative sentences the elements preceding the stressed element cannot have accent. In scientific prose the usual place for accents is in stressed positions, so in reading a paper we usually have to use stress accent.

We use sentences with some word orders more frequently, others less frequently. We say that the stressed element of the sentences with unusual word order has a stronger stress because this way of stressing is more uncommon.

Stressing can be described in terms of more general rules than the rules of accentuation, and it may serve as a basis of the word order rules of the sentences of different types and of typological comparison.

1.5. Our paper deals mainly with the analysis of simple sentences /2/ separating from each other the problems concerning the affirmative and the negative sentences and within these the problems referring to the declarative and the interrogative sentences. Then we proceed to analyze the elliptical sentances /3/ and the sentences with four elements /4/ grouped in the same way as above. The section on prosodic accent /5/ deals with the analysis of the simple affirmative /5.1./ and negative /5.2./ as well as of the elliptic sentences /5.3./. The general questions of stressing and accentuation are attacked in 6. and we touch upon some problems of the relation bew een the word order and context in 7.

2. The word order of simple sentences

2.1. The analysis of the simple sentences can be divided into two parts: the analysis of the affirmative and that of the negative sentences /2.1.1 and 2.2.1/and with respect to both types of sentences we will attack separately the declarative and the interrogative sentences /2.1.2. and 2.2.2./. Primary importance is attached to affirmative declarative sentences. Their analysis serves as the basis for the establishment of the primary rules which undergo considerable changes in the following: they are complemented by new rules or existing rules are deleted.

2.1.1. To analyze the word order of the declarative sentences we want to give first the grammatical rules of the generation of sentences with two or three elements, thereafter the generated elements are permuted and the system obtained in this way is subjected to further analysis /2.1.1.2./.

2.1.1.1. The grammatical rules are used for the generation of the sentences to be examined. They should be considered as fragments of a more complete set of rules which would generate all the Hungarian sentences generatable on this level. An essential shortcoming of our system of rules must be pointed out, namely that it does not differentiate between obligatory and optional complements.

$$S \longrightarrow NP_n + VP$$

$$VP \longrightarrow \begin{cases} V^{intr} + \begin{Bmatrix} \emptyset \\ Adv \\ NP_x \end{Bmatrix} \\ V^{cop} + \begin{Bmatrix} N_n \\ Adj \end{Bmatrix} \\ V^{tr} + NP_a \\ N_n + \begin{Bmatrix} \emptyset \\ NP_x \end{Bmatrix} \\ Adj + \begin{Bmatrix} \emptyset \\ Adv' \\ NP_x \end{Bmatrix} \end{cases}$$

$$NP_i \longrightarrow T + N_i \qquad T \longrightarrow \begin{Bmatrix} a/z/ \\ egy \\ \emptyset \end{Bmatrix} \qquad Adv \longrightarrow \begin{Bmatrix} Adv^{mod} \\ Adv^{loc} \\ Adv^{temp} \end{Bmatrix}$$

$N_n \longrightarrow$ asztal, ház, ember ...

$N_a \longrightarrow$ asztalt, házat, embert ...

$N_x \longrightarrow$ asztalra, asztalon, asztalról, asztalhoz, asztalig, asztaltól, asztalért, asztallal, asztal alatt, asztal felett ...

$V^{intr} \longrightarrow$ megy, úszik, van ...

$V^{cop} \longrightarrow$ volt, lesz ...

$V^{tr} \longrightarrow$ lát, üt, keres ...

$Adj \longrightarrow$ magas, szép, nagy ...

$Adv' \longrightarrow$ nagyon, különösen ...

$Adv^{mod} \longrightarrow$ lassan, gyorsan ...

$Adv \xrightarrow{loc}$ itt, otthon ...

$Adv \xrightarrow{temp}$ ma, holnap ...

By way of explanation let us point to the fact that the symbols x next to N may mean any case or even postposition except for the accusative and nominative. The difference between Adv' and Adv is that the **former** does not contain those adverbs that may be associated only **with verbs.** /Notice, incidentally, that the grammatical rules given above could be simplified if the transitivity of verbs were stored in the dictionary as a grammatical feature associated with the given verb./ The articles are simply enumerated without any classification because no classification would simplify our system of rules. Notice that the articles can be grouped in different ways and if we aimed at establishing only some word order rules we could modify the grouping a/z/, egy, Ø . In this way the definite article could be opposed to the indefinite ones:

$$T \longrightarrow \begin{Bmatrix} \text{Def} \\ \text{Indef} \end{Bmatrix}, \quad \text{Def} \longrightarrow \text{a/z/}, \quad \text{Indef} \longrightarrow \begin{Bmatrix} \text{egy} \\ \emptyset \end{Bmatrix}$$

The lack of article may bring to the fore some generalizations in the use of articles, namely if we oppose the general to the particular we need

$$T \longrightarrow \begin{Bmatrix} \text{Art} \\ \emptyset \end{Bmatrix}, \quad \text{Art} \longrightarrow \begin{Bmatrix} \text{a/ z/} \\ \text{egy} \end{Bmatrix}$$

In the course of the permutation we can denote by the same symbols those elements that behave in the same way, thus the rules can be formulated in a much simpler way. In the case of three elements we work with three symbols: A, B, C which we obtain by means of the following rules:

$$NP_n \longrightarrow A$$

$$V^i \longrightarrow B$$

$$\begin{Bmatrix} N_n \\ Adj \end{Bmatrix} \begin{matrix} \nearrow B & / - V^{cop} / \\ \searrow C & / + V^{cop} / \end{matrix}$$

$$\left\{\begin{matrix} NP_x \\ NP_a \\ \\ Adv \\ Adv' \end{matrix}\right\} \longrightarrow C$$

We give here the reversed system of rules as well.

$$A \longrightarrow NP_n$$

$$B \longrightarrow \left\{\begin{matrix} V^i & \\ N_n & / - V^{cop} / \\ Adj & / - V^{cop} / \end{matrix}\right\}$$

$$C \longrightarrow \left\{\begin{matrix} N_n & / + V^{cop} / \\ Adj & / + V^{cop} / \\ Adv & \\ Adv' & \\ NP_x & \\ NP_a & \end{matrix}\right\}$$

In the course of the analysis of the basic word order we do not consider the adjectives, the adjectival pronouns, numerals, belonging to the group of N /NP/ and the adverbs belonging to the adjective, all these will be tackled within the frame of the word order of NP as pointed out in 1. The word order of the article is fixed but its presence or absence has an influence on the word order. That all this does not belong to the problems connected with the basic word order may be proved by permutation. In Hungarian, which is a language with free word order, we obtain from three elements at least three sequences. On the other hand, from the elements a, vonat, megy, only two sequences can be formed: a vonat megy, megy a vonat, while from vonat, megy, Pestre six.

2.1.1.2. From two elements two different sequences can be formed and each of them contains three variants according to the fact whether the substantive is preceded by a/z/, egy or Ø. In the following we will indicate the stressed element as well or if there is no stress, i.e.if the sentence is neutral, this fact will be indicated by the sign △ . First, we analyze the permutations of the elements

$$/1/ \quad \left\{\begin{matrix} \underline{a} \\ \underline{egy} \\ \emptyset \end{matrix}\right\} \ \underline{\text{vonat megy}}.\text{'} \quad \left\{\begin{matrix} \text{the} \\ \text{a} \\ \emptyset \end{matrix}\right\} \text{train goes'}$$

| | | Stress |
|---|---|---|
| AB: | A vonat megy. | △ |
| | Egy vonat megy. | △ |
| | Vonat megy | A |
| BA: | Megy a vonat. | B |
| | Megy egy vonat. | B |
| | Megy vonat. | B |

In AB the two variants with article are neutral but in the case of the article Ø A is stressed. /In fact, even the first two sentences are not completely neutral, namely B has a slight stress, c.f. Elekfy 334./. The stress occurring in the case of an article Ø is called <u>Ø article stress</u>. In BA the order of two elements was changed which resulted in the stress of B. Such a stress is called <u>permutational stress</u>.

A sentence with two elements may be formed in a way that B stands for an $N_n$ or Adj, for instance, A —— T + $N_n$ B —— Adj, we have also in this case, three variants in each group. The incorrect sentence is denoted by ✗ .

$$/2/ \quad \left\{\begin{matrix} \underline{a} \\ \underline{egy} \\ \emptyset \end{matrix}\right\} \ \underline{\text{ház}} \ \underline{\text{magas}}.\text{'} \left\{\begin{matrix} \text{the} \\ \text{a} \\ \emptyset \end{matrix}\right\} \text{house is high '}$$

| | | Stress |
|---|---|---|
| AB: | A ház magas. | △ |
| | Egy ház magas. | △ |
| | Ház magas. | ✗ |
| BA: | Magas a ház. | B |
| | Magas egy ház | B |
| | Magas ház. | ✗ |

These permutations are characterized by the fact that AB is neutral, in BA, however, B is given a permutational stress. The variants with Ø article are incorrect.

In the case of three elements many sentences may be generated by applying the above rules. It is not necessary to analyze all of them, therefore we restrict ourselves to the analysis of four sentence types which are generated by applying the grammatical rules given above. Notice, however, that some rules are given in detail as $V^{intr} \longrightarrow \left\{ \begin{matrix} V^{intr'} \\ V^{ex} \end{matrix} \right\}$, where the copulative verb is emphasized and in $NP_n$ where N may be a proper noun /$N^{prop}$/ and then the use of the Ø article becomes primary.

/3/ $\left\{ \begin{matrix} \underline{\text{a}} \\ \underline{\text{egy}} \\ \text{Ø} \end{matrix} \right\}$ <u>vonat megy gyorsan /Pestre/</u>. ' $\left\{ \begin{matrix} \text{the} \\ \text{a} \\ \text{Ø} \end{matrix} \right\}$ train goes quickly/to Pest/'

$/ A \longrightarrow T + N_n, \quad B \longrightarrow V^{intr'}, \quad C \longrightarrow \left\{ \begin{matrix} Adv^{mod} \\ N_{re} \end{matrix} \right\} /$

/4/ <u>Péter van otthon</u>. ' Péter is at home'.

$/ A \longrightarrow N_n^{prop}, \quad B \longrightarrow V^{ex}, \quad C \longrightarrow Adv^{loc} /$

/5/ $\left\{ \begin{matrix} \underline{\text{a}} \\ \underline{\text{egy}} \\ \text{Ø} \end{matrix} \right\}$ <u>ház volt magas</u>. ' $\left\{ \begin{matrix} \text{the} \\ \text{a} \\ \text{Ø} \end{matrix} \right\}$ house was high.'

$/ A \longrightarrow T + N_n, \quad B \longrightarrow V^{cop}, \quad C \longrightarrow Adj /$

/6/ <u>Péter olvassa</u> $\left\{ \begin{matrix} \underline{\text{a}} \\ \underline{\text{egy}} \\ \underline{\text{Ø}} \end{matrix} \right\}$ <u>levelet</u>. ' Péter reads $\left\{ \begin{matrix} \text{the} \\ \text{a} \\ \text{Ø} \end{matrix} \right\}$ letter.'

$/A \longrightarrow N_n^{prop}, \quad B \longrightarrow V^{tr}, \quad C \longrightarrow T + N_a /$

In the case of three elements the number of possible permutations amounts to six in contrast to the two in the permutation of two elements:

AB : ABC, ACB
BA : BAC, BCA
and : CAB, CBA

In this enumeration the permutations with two and three elements were placed in a way that the order of AB be the same and the new element stand after them, between them and before them.

Let us examine the stress of sentence /3/.

Here and in the following the order of the permutations will be: ACB, ABC, CBA, CAB, BAC, BCA.

/3/ $\left\{\begin{matrix} \underline{a} \\ \underline{egy} \\ \emptyset \end{matrix}\right\}$ vonat megy gyorsan /Pestre/

| | | Stress |
|---|---|---|
| ACB: | A vonat gyorsan /Pestre/ megy. | C |
| | Egy vonat gyorsan /Pestre/ megy. | C |
| | Vonat gyorsan /Pestre/ megy. | * or /x/ |
| ABC: | A vonat megy gyorsan /Pestre/. | A |
| | Egy vonat megy gyorsan /Pestre/. | A |
| | Vonat megy gyorsan /Pestre/. | A + /Ø stress/ |
| CBA: | Gyorsan /Pestre/ megy a vonat. | C + /per./ |
| | Gyorsan /Pestre/ megy egy vonat. | C + /per./ |
| | Gyorsan /Pestre/ megy vonat. | * or /x/ |
| CAB: | Gyorsan /Pestre/ a vonat megy. | A + /per./ |
| | Gyorsan /Pestre/ egy vonat megy. | A + /per./ |
| | Gyorsan /Pestre/ vonat megy. | $^{/*/}$A + /per./ |
| BAC: | Megy a vonat gyorsan /Pestre/. | $^{/*/}$B |
| | Megy egy vonat gyorsan /Pestre/. | $^{/*/}$B |
| | Megy vonat gyorsan /Pestre/. | B |
| BCA: | Megy gyorsan /Pestre/ a vonat. | $^{/*/}$B |
| | Megy gyorsan /Pestre/ egy vonat. | $^{/*/}$B |
| | Megy gyorsan /Pestre/ vonat. | $^{/*/}$B |

$A^{/*/}$ denotes sentences that are only conditionally correct, i.e. in some contexts. $A^{/x/}$ means that C $\rightarrow$ $N_{re}$ /with the element /<u>Pestre</u>/

is only conditionally correct.

If we have a look at the above permutations we see that in ABC the weak stress falls on A and in ACB on C. Of the two permutations ACB stands nearer to the neutral position. The permutation AB was in the system containing two elements neutral while there is a weak stress in the permutations ACB. ABC, derivable from it; this weak stress is called derivative stress. In BAC and in BCA the permutational stress of BA is preserved on the element B. In CAB and CBA which have no antecedents in the permutations of two elements a stronger stress can be found on the elements A and C, respectively. The Ø article stress increases the stress of A in ABC while in ACB the variant with the Ø article is incorrect if the stressed element is C $\rightarrow$ Adv /gyorsan/ or it results in a conditionally correct sentence if the stressed element is C $\rightarrow N_{re}$ /Pestre/. The same may be observed in CBA.

So, in the permutational system containing six members A and C may have weak or else strong stress while B is always stressed.

$\underline{A}$ : $\underline{A}BC$ $\underline{\underline{A}}$ : $C\underline{\underline{A}}B$

$\underline{C}$ : $A\underline{C}B$ $\underline{\underline{C}}$ : $\underline{\underline{C}}BA$

$\underline{\underline{B}}$ : $\underline{\underline{B}}AC$, $\underline{\underline{B}}CA$

From the two members stressed weakly standing near to the neutral position in ACB the stress of C is weaker than the stress of A in ABC. The permutational stress **of A in CAB** is stronger than that of C in CBA.

If $V^{intr}$ is van 'is' /$V^{ex}$/ then the number of correct sentences may decrease. So for instance

| /4/ | Péter van otthon. | Stress |
|---|---|---|
| ACB : | Péter otthon van. | C |
| ABC : | Péter van otthon. | A |
| CBA : | Otthon van Péter. | C |
| CAB : | Otthon Péter van. | A |
| BAC : | Van Péter otthon. | ~~X~~ |
| BCA : | Van otthon Péter. | ~~X~~ |

As can be seen, of the members formed with verb van in Present BAC and BCA are grammatically incorrect, i.e. those in which the stress should fall on $V^{ex}$. If we use the Past of the verb then we obtain correct sentences.

| | | Stress |
|---|---|---|
| BAC : | Volt Péter otthon. | B |
| BCA : | Volt otthon Péter. | B |

If we analyze a sentence of type /5/ $A \rightarrow T + N_n$, $B \rightarrow V^{cop}$, $C \rightarrow Adj$ : $\left\{\begin{array}{c}\underline{\text{a}}\\ \underline{\text{egy}}\\ \underline{\emptyset}\end{array}\right\}$ <u>ház volt magas</u>, we obtain the following permutations:

| | | Stress |
|---|---|---|
| ACB : | A ház magas volt. | C |
| | Egy ház magas volt. | C |
| | Ház magas volt. | X |
| ABC : | A ház volt magas. | A |
| | Egy ház volt magas. | A |
| | Ház volt magas. | X |
| CBA : | Magas volt a ház. | C |
| | Magas volt egy ház. | C |
| | Magas volt ház. | X |
| CAB : | Magas a ház volt. | A |
| | Magas egy ház volt. | A |
| | Magas ház volt. | X |
| BAC : | Volt a ház magas. | X |
| | Volt egy ház magas. | X |
| | Volt ház magas. | X |
| BCA : | Volt magas a ház. | X |
| | Volt magas egy ház. | X |
| | Volt magas ház. | X |

The stress relations are the same as in the case of sentence /3/ but the sentences BAC and BCA are incorrect in the same way as in the case of the copulative verb, besides the variants with $\emptyset$ article are always incorrect in the same way as in the sentences composed of $N_n$+ Adj. The stress of C is weaker in ACB than that of A in ABC. The features of a sentence of type /5/ can be - as a consequence - derived from the foregoing. If we had $N_n$ instead of Adj we would obtain the same permutations but the members beginning with B would be correct too: Volt Péter katona. Volt katona Péter.

In sentence /6/ $A \rightarrow N_n^{prop}$ because we did not want to increase unnecessarily the number of variants as we had already examined the variants with articles of $N_n$ with respect to sentence /3/; $B \rightarrow V^{tr}$, $C \rightarrow T + N_a$.

Sentence type /6/ Péter olvassa $\left\{\begin{array}{c} \text{a} \\ \text{egy} \\ \emptyset \end{array}\right\}$ levelet

has the following permutations

| | | Stress |
|---|---|---|
| ACB : | Péter a levelet olvassa. | C |
| | Péter egy levelet olvas. | C |
| | Péter levelet olvas. | C |
| ABC : | Péter olvassa a levelet. | A |
| | Péter olvas egy levelet. | A |
| | Péter olvas levelet. | A |
| CBA : | A levelet olvassa Péter. | C |
| | Egy levelet olvas Péter. | C |
| | Levelet olvas Péter. | C |
| CAB : | A levelet Péter olvassa. | A |
| | Egy levelet Péter olvas. | A |
| | Levelet Péter olvas. | A |
| BAC : | Olvassa Péter a levelet. | B |
| | Olvas Péter egy levelet. | B |
| | Olvas Péter levelet. | B |
| BCA : | Olvassa a levelet Péter. | B |
| | Olvas egy levelet Péter. | B |
| | Olvas levelet Péter. | B |

In the permutations the stress falls always on the same element, the stresses can be ordered in the same system as in /3/. In ACB the stress of C is stronger than that of A in ABC which is almost zero. In the variant with Ø article which has a stress on C: Péter levelet olvas. Levelet olvas Péter. what is expressed by the sentence is that the reading of the letter is, as it were, the occupation of Peter as opposed to the example with articles, for instance, with Péter a levelet olvassa. Egy levelet olvas Péter.

To sum up the results obtained from an examination of the sentences we can now establish the rules that produce the sentences with different word orders. The word order rules are entirely separated from the grammatical rules needed for the generation. Each generated sentence must be rearranged after the generation, even in the case when it maintains its word order, namely in the latter case an identical transformation is applied. Though the latter means that we have a superfluous rule but the two systems of rules can be easily separated. The stress can be the result of a permutation only $E$ , but the omission of the article may play a role as well $E_{\emptyset}$ . The latter type of stress is a secondary one, its existence depends on $E$ . The stress may be a weak one, which will be designated by ~ above the stressed member, e.g. $E^{\tilde{A}}$ , and the strong stress by ‾ , e.g. $E^{\bar{A}}$ . The former was called derivative stress, the latter permutational stress. The stress in the case of an Ø article is denoted by ^. The alterations of word order may be conceived as transformations, because the alterations in the order of the symbols A, B, C can only be carried out in the knowledge of the corresponding P-markers. In the following by giving the transformational rules we leave out of account the sentence structure because it is determined by the fact that the symbols A, B, C refer only to sentences with given structure.

To be able to introduce the rules concerning the articles among the transformational rules in an economical way we supply the members A, C with indices before which three different articles may stand.

| i | A |
|---|---|
| 1 $T_1$ + | N |
| 2 $T_2$ + | N |
| 3 Ø + | N |

So, $T \rightarrow a/z/$, $T_2 \rightarrow egy$, $\emptyset \rightarrow \emptyset$ article.

Since the rules referring to the sentences that consist of two elements can be built into those referring to the sentences with three elements only with difficulty, therefore it seems reasonable to separate them.

$$T\ /E^{o}/ : A_iB \quad /I/ \rightarrow \begin{Bmatrix} A_1 \\ A_2 \end{Bmatrix} B \qquad T\ /E_{\emptyset}^{\hat{A}}/ : A_iB\ /I/ \rightarrow \hat{A}_3B$$

$$T\ /E^{\bar{B}}/ : A_iB \rightarrow \bar{B}A_i$$

The first rule refers to a neutral sentence, the second to the sentence with the stress on A with $\emptyset$ article/ A /, and the third to the sentence with a strong permutational stress on B $/\bar{B}/$.

The word order rules of the sentences consisting of three elements may be given - according to the above - as follows:

T /E/

$$T\ /E^{\tilde{A}}/ : A_iBC_j \rightarrow \begin{Bmatrix} \tilde{A}_1 \\ \tilde{A}_2 \end{Bmatrix} BC_j$$

$$T\ /E^{\bar{A}}/ : A_iBC_j \rightarrow C_j \begin{Bmatrix} \bar{A}_1 \\ \bar{A}_2 \end{Bmatrix} B$$

$$T\ /E^{\tilde{C}}/ : A_iBC_j \rightarrow A_i \begin{Bmatrix} \tilde{C}_1 \\ \tilde{C}_2 \end{Bmatrix} B$$

$$T\ /E^{\bar{C}}/ : A_iBC_j \rightarrow \begin{Bmatrix} \bar{C}_1 \\ \bar{C}_2 \end{Bmatrix} BA_i$$

$$T\ /E^{\bar{B}}/ : A_iBC_j \rightarrow \begin{Bmatrix} \bar{B}A_iC_j \\ \bar{B}C_jA_i \end{Bmatrix}$$

T $/E_{\emptyset}/$

$$T\ /E_{\emptyset}^{\hat{A}}/ : A_iBC_j \rightarrow \begin{Bmatrix} \hat{A}_3B \begin{Bmatrix} C_1 \\ C_2 \end{Bmatrix} \\ \begin{Bmatrix} C_1 \\ C_2 \end{Bmatrix} \hat{A}_3B \end{Bmatrix}$$

$$T\ /E_{\emptyset}^{\hat{C}}/ : A_iBC_j \rightarrow \begin{Bmatrix} \begin{Bmatrix} A_1 \\ A_2 \end{Bmatrix} \hat{C}_3\ B \\ \hat{C}_3\ B \begin{Bmatrix} A_1 \\ A_2 \end{Bmatrix} \end{Bmatrix}$$

In the transformation $E_{\emptyset}$ we do not think it would be necessary to make reference to the permutation in which $E_{\emptyset}$ occurs because this does not result in a perceivable difference in stressing.

These general rules, however, are in need of further refinements. So from the group of the rules /E/ it is impossible to arrive at

$T\ /E^{\bar{B}}/$ if $B \rightarrow \begin{Bmatrix} V^{cop} \\ V^{ex} \end{Bmatrix}$ and $V^{ex} \rightarrow$ *van*; if $B \rightarrow \begin{Bmatrix} Adj \\ N \end{Bmatrix}$ then $T/E^{\bar{B}}/$:

$A_iB \rightarrow \bar{B}\begin{Bmatrix} A_1 \\ A_2 \end{Bmatrix}$. In the group of the rules $/E_{\emptyset}/$, if $B \rightarrow \begin{Bmatrix} Adj \\ N \end{Bmatrix}$ then $/E_{\emptyset}^{\hat{A}}/$ is not possible. If $C \rightarrow Adv$, then C can obtain stress only by the following rule $T\ /E^{C}/ : A_iBC \rightarrow \begin{Bmatrix} \begin{Bmatrix} A_1 \\ A_2 \end{Bmatrix} CB \\ CB\begin{Bmatrix} A_1 \\ A_2 \end{Bmatrix} \end{Bmatrix}$.

2.1.2. Only those were analyzed of the *interrogative sentences* that are formed by the alteration of the intonation. The data concerning stress obtained from the sentences refer equally to declarative and interrogative sentences because in the latter we have the same stress conditions as in the former. As a consequence, we have the same transformational rules with respect to the interrogative and declarative sentences, so they need not be analyzed separately. The only difference is given by the question transformation: $T_Q$ : ABC $\rightarrow$ ABC ?

2.2. In order to analyze the regularities of stressing in *negative sentences* we give in Appendix 1. the table containing the word order relations in sentences with two elements /1/ and /2/ where the words are replaced by symbols. From the sentences with three elements only sentence /6/ was analyzed /cf. Appendix 2./. The cases of double negation are not given in the table, because nBnA is the same as nA. nBnC the same as nC and nAnC yields incorrect sentences.

2.2.1. Of the negative sentences the *declarative sentences* first will be given a full analysis in Appendix 1. The aim of this investigation is to find out in what the word order of the negative sentences differs from that of the declarative sentences.

The grammatical rules needed for the generation are the same in the case of negative sentences as for declarative sentences but the particle $n \rightarrow$ *nem* is introduced before the elements it refers to.

In the sentences with two elements if we negate B, then $T_{nB}$ : AB $\rightarrow$ AnB is the initial sentence. In the variants with *a/z/* and egy articles of sentences with the word order AnB the negative particle of B gets weak stress and to this extent it differs from the affirmative sentences which are neutral. Such stress is called negation stress. In the nAB and nAnB sentences there is a negation stress on A, the BnA and nBnA members are incorrect in all their variants.

Sentence /2/ with nominal predicate /B $\rightarrow$ Adj/ reveals the same properties as /1/ in this respect. However, the members with an $\emptyset$ article are incorrect as in the case of declarative sentences.

In the sentence with three elements /6/ if we negate B, then in the permutations $T_{nB}$:ABC $\rightarrow$ AnBC the negation stress falls on B in contrast to the declarative sentences though the stresses of B are not always the same, namely it is weaker in the members CAB and CBA than in ABC and ACB.

If we negate A:$T_{nA}$:ABC $\rightarrow$ nABC, then the negation stress can only fall on A in the members nABC and CnAB.

If we negate C:$T_{nC}$:ABC $\rightarrow$ ABnC, then the negation stress can only fall on C in the member AnCB and nCBA.

In the case of double negation $T_{nBnA}$:ABC $\rightarrow$ nAnBC and $T_{nBnC}$:ABC $\rightarrow$ AnBnC we have the same situation as when A or C is negated.

The variants with the article egy yield unacceptable sentences.

On the basis of what has been said we can formulate the following rules:

Sentences with two elements

$T_{nB}$ : AB $\rightarrow$ AnB

T /E/

$T_{nB}$ /$E^{\tilde{B}}$/ : $A_i$nB /I/ $\rightarrow$ $A_i$n$\tilde{B}$

$T_{nB}$ /$E^{\bar{B}}$/ : $A_i$nB $\rightarrow$ n$\bar{B}A_i$

$T_{nA}$ : AB $\rightarrow$ nAB

T /E/

$T_{nA}$ /$E^{\tilde{A}}$ / : n$A_i$B $\rightarrow$ n$\begin{Bmatrix}\tilde{A}_1\\ \tilde{A}_2\end{Bmatrix}$B

T /$E_{\emptyset}$/

$T_{nA}$/$E^{\hat{A}}_{\emptyset}$/ : n$A_i$B $\rightarrow$ n$\hat{A}_3$B

$T_{nAnB}$ : AB $\rightarrow$ nAnB

T /E/

$T_{nAnB}$ /$E^{\tilde{A}}$/ : n$A_i$nB $\rightarrow$ n$\begin{Bmatrix}\tilde{A}_1\\ \tilde{A}_2\end{Bmatrix}$nB

T /$E_{\emptyset}$/

$T_{nAnB}$ /$E^{\hat{A}}_{\emptyset}$/ : n$A_i$nB $\rightarrow$ n$\hat{A}_3$nB

/I/ denotes the identical transformation.

Notice, if $B \rightarrow \left\{\begin{matrix} Adj \\ N_n \end{matrix}\right\}$ then there is no $T/E_{\emptyset}$ /.

Sentences with three elements.

A does not receive an index because: $A \rightarrow N^{prop}$ , $N^{prop} \rightarrow$ <u>Péter</u>.

$T_{nB}$ : $ABC \rightarrow AnBC$

T/E/

$$T_{nB}\ /E^{\tilde{B}}/: AnBC_j \rightarrow \left\{\begin{matrix} A\left\{\begin{matrix} C_1 \\ C_3 \end{matrix}\right\} n\tilde{B} \\ An\tilde{B}\left\{\begin{matrix} C_1 \\ C_3 \end{matrix}\right\} \\ \left\{\begin{matrix} C_1 \\ C_3 \end{matrix}\right\} n\tilde{B}A \\ \left\{\begin{matrix} C_1 \\ C_3 \end{matrix}\right\} An\tilde{B} \end{matrix}\right\}$$

$$T_{nB}\ /E^{\bar{B}}/: AnBC_j \rightarrow \left\{\begin{matrix} n\bar{B}A\left\{\begin{matrix} C_1 \\ C_3 \end{matrix}\right\} \\ n\bar{B}\left\{\begin{matrix} C_1 \\ C_3 \end{matrix}\right\} A \end{matrix}\right\}$$

Notice that A has not been differentiated according to the articles, thus $T/E^{\hat{A}}_{\emptyset}$ / could not be applied.

$T_{nA}$ : $ABC \rightarrow nABC$

T /E/

$$T_{nA}\ /E^{\tilde{A}}/: nABC_j \rightarrow n\tilde{A}B\left\{\begin{matrix} C_1 \\ C_3 \end{matrix}\right\}$$

$$T_{nA}\ /E^{\bar{A}}/: nABC_j \rightarrow \left\{\begin{matrix} C_1 \\ C_3 \end{matrix}\right\} n\bar{A}B$$

The same remark as above.

$T_{nC}$: ABC $\rightarrow$ ABnC

T /E/ | T /$E_\emptyset$/

$T_{nC}$ /$E^{\bar{C}}$/: $ABnC_j \rightarrow An\bar{C}_1B$

$T_{nC}$ /$E^{\bar{C}}$/: $ABnC_j \rightarrow n\bar{C}_1BA$

$T_{nC}$/$E_\emptyset^{\hat{C}}$/: $ABnC_j \rightarrow \begin{Bmatrix} An\hat{C}_3B \\ n\hat{C}_3BA \end{Bmatrix}$

$T_{nAnB}$: ABC $\rightarrow$ nAnBC  like $T_{nA}$
$T_{nCnB}$: ABC $\rightarrow$ AnBnC  like $T_{nC}$

2.2.2. We have the same transformational rules with respect to the interrogative sentences.

2.3.1. In the foregoing we have analyzed the permutations of affirmative and negative simple sentences and have stated which sentences are correct and which element in them bears the stress and in accordance with our observations we have established the word order rules. In the course of our investigations we have taken the stress of the sentence for granted and have not tried to explain why the stress fell on this or that member and not on some other. Further, we shall try to explain why those sentences are incorrect that were found to be so by the informants.

Let us briefly enumerate the factors that lead to a given stress. From the neutral sentence A vonat megy, by omitting the article we obtain an $\emptyset$ article stress: Vonat megy./$e_\emptyset$/. If we transpose the elements of the sentence: Megy a vonat., there is a permutational stress /$e_p$/. In the case when the verb is negated: A vonat nem megy. there is a weaker negation stress /$e_{n-V}$/, when the substantive is negated: Nem a vonat megy.there is a stronger negation stress /$e_{n-N}$/. The question may be raised what happens when there is a possibility to lay the stress on different elements and thus the stresses could be constrasted. Let us examine the variants with a and $\emptyset$ articles of the permutations of /1/. For the sake of comparison we indicate the difference in the stresses by numbers, so for instance, $e_\emptyset = 2$, $e_{n-V} = 2.5$, $e_p = 3$, $e_{n-N} = 3.5$

Stress

| | | |
|---|---|---|
| AB: | A vonat megy. | |
| | Vonat megy. | $A(e_{\emptyset})$ |
| BA: | Megy a vonat. | $B(e_p)$ |
| | Megy vonat. | $B(e_p = 3)$: $A(e_{\emptyset} = 2) > B$ |
| AnB: | A vonat nem megy | $B(e_{n-V})$ |
| | Vonat nem megy. | $A(e_{\emptyset} = 2)$: $B(e_{n-V} = 2.5) > B$ |
| nBA: | Nem megy a vonat. | $B(e_p + e_{n-V})$ |
| | Nem megy vonat. | $B/(e_p = 3)+(e_{n-V} = 2.5)= 5.5/$: $A(e_{\emptyset}=2) > B$ |
| nAB: | Nem a vonat megy. | $A(e_{n-N})$ |
| | Nem vonat megy. | $A(e_{n-N} + e_{\emptyset})$ |
| BnA: | *Megy nem a vonat. | $B(e_p = 3)$: $A(e_{n-N} = 3.5) > *$ |
| | *Megy nem vonat. | $B(e_p = 3)$ :$A/(e_{n-N} = 3.5)+(e_{\emptyset}=2) = 5.5/ > *$ |
| nAnB: | Nem a hajó nem úszik. | $A(e_{n-N} = 3.5)$: $B(e_{n-V} = 2.5) > A$ |
| | Nem hajó nem úszik. | $A/(e_{n-N} = 3.5)+(e_{\emptyset} = 1)= 4.5/$: |
| | | $B(e_{n-V} = 2.5) > A$ |
| nBn: | *Nem úszik nem a hajó. | $B/(e_p = 3)+(e_{n-V} = 2.5)= 5.5/$:$A(e_{n-N} = 3.5) > *$ |
| | *Nem úszik nem hajó. | $B/(e_p = 3)+(e_{n-V} = 2.5)= 5.5/$: |
| | | $A/(e_{n-N} = 3.5)+(e_{\emptyset} = 2)= 5.5/ > *$ |

It is easy to see which stress is the strongest among the possible stresses and which element receives the stress of the sentence. If two strong stresses come up against each other, then the sentence will be incorrect. If two weak stresses come together, it seems reasonable to apply prosodic accent to be able to decide about the stress. It should be noted, however, that the numbers indicate only tentatively the strength of the stresses.

Sentences with three elements can also be analyzed this way but the weak derivative stress has to be taken into consideration as well $/e_d=0.5/$ which may have different strength in the two basic members. Let us examine some interesting cases in the following sentence

/6/ A tanuló a levelet olvassa.

ACB: A tanuló a levelet olvassa. $C(e_d)$

A tanuló levelet olvas. $C(e_d + e_\emptyset)$

*Tanuló levelet olvas. $A(e_\emptyset=2) : C/(e_d=0.5)+(e_\emptyset=2)=2.5/ >$ *

ABC: A tanuló olvassa a levelet. $A(e_d)$

AnBC: A tanuló nem olvassa a levelet. $A(e_d=0.5) : B(e_{n-V}=2.5) >P$

So we can explain why those sentences are incorrect in which there are two substantives with Ø articles and why the stress is moved to the verb with a negative particle. We do not obtain, however, a satisfactory solution in such cases as

ACB : A vonat gyorsan megy. $C(e_d)$

Vonat gyorsan megy. $A(e_\emptyset=2) : C(e_d=0.5) >$ *

Though the stress of A is much stronger, it is, in fact, a Ø article stress, that of the adverb, on the other hand, is only a derivative stress, yet the sentence is incorrect. Obviously, in such cases we have to attribute a stronger stress to the adverb as in the sentence * Tanuló levelet ír.

2.3.2. This analysis, however, does not offer an explanation for cases when the sentence is incorrect because of the fact that some part-of-speech subgroups /e.g. verb van, copula/ can scarcely ever receive any stress. Besides, some words may have special word order properties which must be taken into consideration. It would be very desirable that the words should be provided alongside grammatical information also with some indication as to word order features which would be of negative character, i.e. they would inhibit certain word order positions.[5] This question, however, needs further and separate study, but before many interesting questions must be clarified which would also require the results of formal semantic theory.

3. The elliptical sentences

3.1. Here,too,let us begin with the affirmative sentence. First we have again to delimit the scope of our investigations. Since in the case of the ellipsis of sentences with two elements only a single element is obtained, only sentences with three elements must be considered. From these only such sentences are analyzed in which there is no $NP_n$. For if we ommit the verb from a sentence as

Péter olvas/sa/ $\left\{\begin{matrix} a \\ egy \\ \emptyset \end{matrix}\right\}$ levelet. we obtain a sentence that is only correct if either element of it is uttered with prosodic accent and if we make a pause between the two elements. E.g. Péter # levelet. However, we do not want to deal with this type of sentence in the present work. If we ommit the complement of the verb in the case of sentences with three elements: Péter olvassa. Péter olvas. we obtain sentences with two elements. The second type has already been analyzed, the first has to be uttered with prosodic accent. So what remains is the ellipsis of $NP_n$.

3.1.1.1. The ellipsis has to be carried out by means of an elliptic transformation: in the case of an A member: $T^{Ell/A/}$ : ABC $\rightarrow$ BC. The word order transformations can only be carried out on the output of the former transformation. From the sentences with three members we examine first the ellipsis of the sentence

/6/ Péter olvas/sa/ $\left\{\begin{matrix} a \\ egy \\ \emptyset \end{matrix}\right\}$ levelet. namely /6 ell/ Olvassa $\left\{\begin{matrix} a \\ egy \\ \emptyset \end{matrix}\right\}$ levelet.

in Appendix 3. These three sentences are analyzed as affirmative and negative sentences and then within each group as declarative and interrogative sentences. We have chosen this sentence type because we want to show the behaviour of the article, too, which would not always be possible in the case of the ellipsis of sentence

/3/ $\left\{\begin{matrix} a \\ egy \\ \emptyset \end{matrix}\right\}$ vonat megy gyorsan /Pestre/. In the case of the ellipsis of sentence /6/ we obtain sentences that behave like /6 ell/

Megyünk $\left\{\begin{matrix} az \\ egy \\ \emptyset \end{matrix}\right\}$ iskolába. which can be considered as an elliptical variant of /3/ with the difference that we have instead of Pestre a substantive with article. Sentence /4 ell/ Van otthon. from /4/ Péter van otthon will be considered, too.

3.1.1.2. On the basis of the table it is apparent that the stress falls on B in the member BC and on C in the member CB. In sentence /4 ell/ the member BC only gives a correct sentence if it is uttered with a prosodic accent. The variant with an Ø article yields definitely correct sentences in the member CB but in the member BC only under certain contextual restrictions.

Let us now formulate the rules for the ellipsis. The first rule is the elliptical transformation referring to the member A:

$$T^{Ell\ /A/} : ABC \longrightarrow BC$$

The emphatic transformations and rules for the forming of ø article stress:

$T\ /E/$ $\qquad\qquad$ $T\ /E_{\emptyset}/$

$$T^{Ell/A/}\ /E^{\tilde{B}}/\ :\ BCj\ /I/ \rightarrow \tilde{B}Cj$$

/except for $B \rightarrow V^{ex}$ /

$$T^{Ell/A/}\ /E^{\tilde{C}}/\ :\ BCj \rightarrow \begin{Bmatrix} \tilde{C}_1 \\ \tilde{C}_2 \end{Bmatrix} B \qquad\qquad T^{Ell/A/}\ /E_{\emptyset}^{\hat{C}}/\ :\ BCj \rightarrow \hat{C}_3B$$

3.1.2. The transformational rules for interrogative sentences are the same as for declarative ones but the question transformation has to be introduced: $T_Q\ :\ BC \rightarrow BC?$

3.2.1. As to the negative declarative sentences we focus our attention first on nB. Here in either member B bears the stress but the variant with the article egy is not unacceptable either. If C is negated /nC/ and in the case of double negation /nBnC/ each variant of BC is incorrect, in the member CB C is stressed but the variant with the article egy does not yield correct sentences.

All that has been said can be formulated in the following rules. In the elliptical declarative sentence the negation of B:

$$T_{nB}^{Ell/A/}\ :\ BC \rightarrow nBC$$

Then the following emphatic transformations can be carried out:

$T\ /E/$

$$T_{nB}^{Ell/A/}\ /E^{\tilde{B}}/\ :\ nBCj \rightarrow \begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix} n\tilde{B}$$

$$T_{nB}^{Ell/A/}\ /E^{\bar{B}}/\ :\ nBCj\ /I/ \rightarrow n\bar{B} \begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix}$$

The negation of C in the elliptical declarative sentence:

$$T_{nC}^{Ell/A/}\ :\ BC \rightarrow BnC$$

Here the following transformation can be carried out:

$$T\ /E/ \qquad\qquad\qquad\qquad T\ /E_{\emptyset}/$$

$$T^{Ell/A/}_{nC}\ /E^{\tilde{C}}/ : BnCj \rightarrow n\tilde{C}_1B, \qquad T^{Ell/A/}_{nC}\ /E^{\hat{C}}_{\emptyset}/ : BnC \rightarrow n\hat{C}_3I$$

In the case of double negation

$$T^{Ell/A/}_{nBnC}: BC \rightarrow nBnC$$

the emphatic transformations are the same

$$T\ /E/ \qquad\qquad\qquad\qquad T/E_{\emptyset}/$$

$$T^{Ell/A/}_{nBnC}\ /E^{\tilde{C}}/ : nBnCj \rightarrow n\tilde{C}_1nB \qquad T^{Ell/A/}_{nBnC}\ /E^{\hat{C}}_{\emptyset}/ : nBnC \rightarrow N\hat{C}_3nB$$

3.2.2. The stress relations and the transformational rules of the <u>interrogative sentences</u> are the same as those of the declarative sentences.

4. The word order of sentences with four elements.

4.1. We want to deal only with some word order problems in the case of sentences with four elements because we do not aim at a detailed discussion of this type of sentence. What we want to find out about is the extent to which the regularities established so far are valid in this case and by what rules our system has to be complemented.

4.1.1. We treat only the <u>affirmative sentences</u> and only one sentence of them which can be considered as the most characteristic of this type, namely the sentence

/7/ <u>Péter olvassa a levelet gyorsan</u>.

$/A \rightarrow N^{prop}_{n},\ B \rightarrow V^{tr},\ C \rightarrow a{+}N_a,\ D \rightarrow Adv^{\ mod}/$

'Péter reads the letter quickly.'

Sentence /7/ consists of sentence /6/ with three elements to which the adverb of sentence /3/ is added. So its generation requires the same rules as /6/ but $N_a$ is followed by an Adv. If the same verb were intransitive and there were no object, then we would obtain a sentence of type /3/.

The permutations are ordered according to the permutations of the tree-element sentence: the first row begins with A, the second with C, the third with B and the last with D.

| | | Stress |
|---|---|---|
| ACBD: | Péter a levelet olvassa gyorsan | C |
| ACDB: | Péter a levelet gyorsan olvassa. | D |
| ABCD: | Péter olvassa a levelet gyorsan. | A |
| ABDC: | Péter olvassa gyorsan a levelet. | A |
| ADCB: | Péter gyorsan a levelet olvassa. | C |
| ADBC: | Péter gyorsan olvassa a levelet. | D |
| | | |
| CBAD: | A levelet olvassa gyorsan Péter. | C |
| CBDA: | A levelet olvassa Péter gyorsan. | C |
| CABD: | A levelet Péter olvassa gyorsan. | A |
| CADB: | A levelet Péter gyorsan olvassa. | D |
| CDAB: | A levelet gyorsan Péter olvassa. | A |
| CDBA: | A levelet gyorsan olvassa Péter. | D |
| | | |
| BACD: | Olvassa Péter a levelet gyorsan. | B |
| BADC: | Olvassa Péter gyorsan a levelet. | B |
| BCAD: | Olvassa a levelet Péter gyorsan. | B |
| BCDA: | Olvassa a levelet gyorsan Péter. | B |
| BDAC: | Olvassa gyorsan Péter a levelet. | B |
| BDCA: | Olvassa gyorsan a levelet Péter. | B |
| | | |
| DABC: | Gyorsan Péter olvassa a levelet. | A |
| DACB: | Gyorsan Péter a levelet olvassa. | C |
| DBAC: | Gyorsan olvassa Péter a levelet. | D |
| DBCA: | Gyorsan olvassa a levelet Péter. | D |
| DCAB: | Gyorsan a levelet Péter olvassa. | A |
| DCBA: | Gyorsan a levelet olvassa Péter. | C |

In the 24 members every element may be stressed. Stressed B stands always in the first place, A,C,D in the first, second, and third place. The stressed B indicates continuous action. A has weaker stress at the beginning of the sentence but a stronger stress in the second and third place. C and D receive weaker stress in the second place, but stronger stress in the first and third place. All that has been said about the stresses of A,B,C coincides with our observations about the sentences with three elements. The stress conditions of D coincide with those of C and this is quite understandable because both are complements of the verb and as a consequence both may be C in the sentence with three elements. Yet a slight difference may be perceived between the stresses of the two elements. If C and D stand in the second place with weaker stress, then there is some difference in the strength of the stress: in the sentence ACBD: Péter a levelet olvassa gyorsan. the word levelet is more heavily stressed than gyorsan in the same position: Péter gyorsan olvassa a levelet. Much work must still be done in this direction contrasting the different complements of the verb.

We have not analyzed sentences with the copulative verb. The sentence

/8/ Péter van otthon ma.

'Péter is at home today'.

$/A \rightarrow N_n^{prop}$ , $B \rightarrow V^{ex}$, $C \rightarrow Adv^{loc}$ , $D \rightarrow Adv^{temp}$ /

is formed from /4/ by addition of a complement. Such sentences, however, cannot be divided into two sentences, because though we have Péter van otthon, we have not $^{x}$Péter van ma. Here we obtain incorrect sentences when a member begins with B, otherwise the same stress conditions are obtained as in /7/. In the weakly stressed second position: ACBD: Péter otthon van ma. and ADBC: Péter ma van otthon. it seems that the former has a weaker stress and is therefore nearest to the neutral sentence. Let us also mention the copula type sentence with four elements as, for instance, the sentence

/9/ A ház volt új tavaly.

$/A \rightarrow a + N_n$, $B \rightarrow V^{cop}$ , $C \rightarrow Adj$, $D \rightarrow Adv^{temp}$ /

'The house was new last year'

that can be derived from /5/ by complementation. This sentence cannot be divided into two sentences, because: A ház volt új. is a correct sentence but $^{x}$ A ház volt tavaly. not. From the permutations those beginning with B are incorrect, otherwise the stress conditions are the same as in type /7/. If C and D stand in the second place of the sentence: ACBD: **A ház új volt tavaly.** ADBC: A ház tavaly volt új., that sentence is given weaker stress which contains C in the second place.

4.1.2. Though our aim was not to analyze the sentence with four elements in full detail, yet it turned out that the stressing rules are the same for the sentences with four elements as for those with three elements, but some particular features of the former must be determined in the word order of the complements of the verb. Let us try to state shortly the word order rules of the sentence of type /7/. Since we have only examined the variant with the article a, we have no rules for the ∅ article stress. The nominal elements are not provided with indices since we have neglected the question of nominal variants. Now we have the following rules:

T/E/

$$T\ /E^{\tilde{A}}/:\ ABCD \longrightarrow \left\{\begin{matrix} \tilde{A}BCD \\ \tilde{A}BDC \end{matrix}\right\}$$

$$T\ /E^{\bar{A}}/:\ ABCD \longrightarrow \left\{\begin{matrix} C\bar{A}BD \\ D\bar{A}BC \\ CD\bar{A}B \\ DC\bar{A}B \end{matrix}\right\}$$

$$T\ /E^{\bar{B}}/:\ ABCD \longrightarrow \left\{\begin{matrix} \bar{B}ACD \\ \bar{B}ADC \\ \bar{B}CAD \\ \bar{B}CDA \\ \bar{B}DAC \\ \bar{B}DCA \end{matrix}\right\}$$

$$T\ /E^{\tilde{C}}/:\ ABCD \longrightarrow \left\{\begin{matrix} A\tilde{C}BD \\ D\tilde{C}BA \end{matrix}\right\}$$

$$T\ /E^{\bar{C}}/:\ ABCD \longrightarrow \left\{\begin{matrix} \bar{C}BAD \\ \bar{C}BDA \\ AD\bar{C}B \\ DA\bar{C}B \end{matrix}\right\}$$

$$T\ /E^{\tilde{D}}/:\ ABCD \longrightarrow \left\{\begin{matrix} A\tilde{D}CB \\ C\tilde{D}BA \end{matrix}\right\}$$

$$T\ /E^{\bar{D}}/:\ ABCD \rightarrow \begin{Bmatrix} \bar{D}BAC \\ \bar{D}BCA \\ AC\bar{D}B \\ CA\bar{D}B \end{Bmatrix}$$

To differentiate the members according to the strength of the stress would be very difficult, so we do not enter into this question here. The existence of forms with identical stress is justified by the fact that otherwise we could not place appropriately our sentences in the different contexts. Such a classification of the sentences, however, belongs already to the scope of pragmatic investigation.

4.2. We do not examine the word order problems of sentences with four elements further but in conclusion let us allude to the elliptical sentences derived from them. Our remarks will refer only to the ellipsis of the element A. We obtain from the sentence ABCD by applying a transformational rule BCD:

$$T^{Ell/A/}\ :\ ABCD \rightarrow BCD$$

Instead of /7/ we examine therefore the following sentence:

/7 ell/ : Olvassa a levelet gyorsan.
'Reads the letter quickly'.

We give below the six permutations in the usual way:

| | | Stress |
|---|---|---|
| CBD: | A levelet olvassa gyorsan. | C |
| CDB: | A levelet gyorsan olvassa. | D |
| BCD: | Olvassa a levelet gyorsan. | B |
| BDC: | Olvassa gyorsan a levelet. | B |
| DBC: | Gyorsan olvassa a levelet. | D |
| DCB: | Gyorsan a levelet olvassa. | C |

As to the stress we see that B is stressed in the first place of the sentence, C and D in the first and second places. It is rather difficult to form a judgement of the strength of the stress. The stress of D is weaker in the member DBC and stronger in CBD. On the other hand, C is hardly less strongly stressed in CDB than in DCB. The neutral sentence is best approximated by DBC: Gyorsan olvassa a levelet. To clarify this problem one has to take into consideration that the elements standing in the second place in the sentence with four elements do not get a weak stress and that the element standing in the third place is stronger than the former. Omitting A the originally second element is shifted to the first place and becomes stressed, and on the other hand, the element standing originally in the third place comes to the second place and so it loses something of its stress. Therefore it is very difficult to differentiate between the two stressed members of C. All that has been said so far can be formulated by the following transformational rules:

T /E/

$$T^{Ell/A/}\ /E^{\bar{B}}/:\ BCD \rightarrow \left\{\begin{matrix} \bar{B}CD \\ \bar{B}DC \end{matrix}\right\}$$

$$T^{Ell/A/}\ /E^{\tilde{C}}/:\ BCD \rightarrow \tilde{C}BD$$

$$T^{Ell/A/}\ /E^{\bar{C}}/:\ BCD \rightarrow D\bar{C}B$$

$$T^{Ell/A/}\ /E^{\tilde{D}}/:\ BCD \rightarrow \tilde{D}BC$$

$$T^{Ell/A/}\ /E^{\bar{D}}/:\ BCD \rightarrow C\bar{D}B$$

5. Questions of the prosodic accent

5.1. Here,too, we focus our attention on the analysis of affirmative sentences. At the beginning, let us again delimit the scope of our investigations. By prosodic accent we understand the prominence given to a word of the sentence by acoustic means. The prosodic accent /P/ may fall on one word of the sentence or on two words at the same time. If P falls on two elements of the sentence, then there is also a pause in the sentence designated by #. We deal only with the case when there is one P in the sentence. In the following we analyze our sentences and on the basis of the analysis we state the rules of accentuation. We restrict ourselves, however, to the most important types. In the present paper only a portion of the data collected for the monograph are presented. The accent is designated by $\underline{\underline{\qquad}}$ if P is primary and a stress accent which means that the place of the accent is identical with that of the stress and by $\underline{\qquad}$ if P falls on a word that is not stressed, i.e. if P is secondary.

5.1.1. We begin the analysis of the declarative sentences with the sentences with two elements. The elements of sentence types /1/ and /2/ may be accentuated in the following way /if only one element is accentuated/:

/1/ $\left\{\begin{matrix}\text{a}\\ \text{egy}\\ \emptyset\end{matrix}\right\}$ vonat megy. '$\left\{\begin{matrix}\text{the}\\ \text{a}\\ \emptyset\end{matrix}\right\}$ train goes.' Accent

| AB | BA | AB | | BA |
|---|---|---|---|---|
| A vonat megy. | Megy a vonat. | A | B | B |
| Egy vonat megy. | Megy egy vonat. | A | B | B |
| Vonat megy. | Megy vonat. | A | | B |

/2/ $\left\{\begin{matrix}\text{a}\\ \text{egy}\\ \emptyset\end{matrix}\right\}$ ház magas.' $\left\{\begin{matrix}\text{the}\\ \text{a}\\ \emptyset\end{matrix}\right\}$ house is high.' Accent

| AB | BA | AB | | BA |
|---|---|---|---|---|
| A ház magas. | Magas a ház. | A | B | B |
| Egy ház magas. | Magas egy ház. | A | B | ×B |
| Ház magas. | Magas ház | | × | × |

In the variants with the articles a and egy of the member AB either element may have stress accent, in the variant with ∅ article only A, more precisely, only in /1/, because in /2/ this variant is incorrect. In the BA member it may fall only on B. By comparing the accent with the stress /2.1.1.2./ we see that in the neutral sentences /a and egy variants of AB/ either element may be eligible for P, in the other sentences the stressed and the stress accentuated elements coincide.

The accentuation of the sentences with three elements are exemplified by sentence /6/. This will be compared with a sentence related to /4/ in which Péter is replaced by article + substantive,

/6/ Péter olvas/sa/ $\left\{\begin{matrix}\text{a}\\ \text{egy}\\ \emptyset\end{matrix}\right\}$ levelet. 'Peter reads $\left\{\begin{matrix}\text{the}\\ \text{a}\\ \emptyset\end{matrix}\right\}$ letter'.

/4'/ $\left\{\begin{matrix}\text{a}\\ \text{egy}\\ \emptyset\end{matrix}\right\}$ tanár van bent. '$\left\{\begin{matrix}\text{the}\\ \text{a}\\ \emptyset\end{matrix}\right\}$ teacher is indoors' .

| ACB /6/ | /4'/ | Accent /6/ | /4'/ |
|---|---|---|---|
| Péter a levelet olvassa. | A tanár bent van. | CB | C |
| Péter egy levelet olvas. | Egy tanár bent van. | CB | C |

| | | | |
|---|---|---|---|
| Péter levelet olvas. | Tanár bent van. | CB | /x/C |
| ABC | | | |
| Péter olvassa a levelet. | A tanár van bent. | ABC | AC |
| Péter olvas egy levelet. | Egy tanár van bent. | ABC | AC |
| Péter olvas levelet. | Tanár van bent. | ABC | AC |
| CBA | | | |
| A levelet olvassa Péter. | Bent van a tanár. | CBA | CA |
| Egy levelet olvas Péter. | Bent van egy tanár. | CBA | CA |
| Levelet olvas Péter. | Bent van tanár. | CBA | /x/CA |
| CAB | | | |
| A levelet Péter olvassa. | Bent a tanár van. | A B | A |
| Egy levelet Péter olvas. | Bent egy tanár van. | A B | A |
| Levelet Péter olvas. | Bent tanár van. | A B | A |
| BAC | | | |
| Olvassa Péter a levelet. | Van a tanár bent. | B C | x |
| Olvas Péter egy levelet. | Van egy tanár bent. | B C | B |
| Olvas Péter levelet. | Van tanár bent. | B C | B |
| BCA | | | |
| Olvassa a levelet Péter. | Van bent a tanár. | B A | x |
| Olvas egy levelet Péter. | Van bent egy tanár. | B A | B |
| Olvas levelet Péter. | Van bent tanár. | B A | B |

In /6/ the originally stressed element has the accent on it but also the following element may be accentuated but in the latter case a complementary sentence is required, so, for instance, Péter a levelet olvassa /és nem irja/. Such an accent is called secondary. A and C may get secondary accent as well, e.g. Péter olvassa a levelet/és nem a lapot/. which seems, however, less common than the accentuation of B.

In /4'/ there is a copulative verb which cannot receive secondary accent, and stress accent only in the case if the substantive is preceded by the article egy or Ø. If B is accentuated then the following element cannot be accented.

On the basis of the foregoing we are now able to state the rules of accentuation. In the case of two elements

$$T\ /P^{\underline{\underline{A}}}/:\ A_iB \longrightarrow \underline{\underline{A_i}}B$$

$$T\ /P^{\underline{\underline{B}}}/:\ A_iB \longrightarrow \left\{\begin{matrix} \left\{\begin{matrix} A_1 \\ A_2 \end{matrix}\right\} \underline{\underline{B}} \\ \underline{\underline{B}}A_i \end{matrix}\right\}$$

If $B \longrightarrow \left\{\begin{matrix} Adj \\ N \end{matrix}\right\}$, instead of $A_i$ we have $\left\{\begin{matrix} A_1 \\ A_2 \end{matrix}\right\}$.

In the case of three elements

$$T/P^{\underline{\underline{A}}}/:\ ABC \longrightarrow \left\{\begin{matrix} \underline{\underline{A}}BC \\ C\underline{\underline{A}}B \end{matrix}\right\}$$

$$T\ /P^{\underline{\underline{B}}}/:\ ABC \longrightarrow \left\{\begin{matrix} \underline{\underline{B}}AC \\ \underline{\underline{B}}CA \end{matrix}\right\}$$

$$T\ /P^{\underline{A}}/:\ ABC \longrightarrow \left\{\begin{matrix} CB\underline{A} \\ BC\underline{A} \end{matrix}\right\}$$

$$T\ /P^{\underline{B}}/:\ ABC \longrightarrow \left\{\begin{matrix} AC\underline{B} \\ A\underline{B}C \\ C\underline{B}A \\ CA\underline{B} \end{matrix}\right\}$$

if: $B \longrightarrow V^{ex}$, $T\ /P^{\underline{\underline{B}}}/:\ A_iBC \left\{\begin{matrix} \underline{\underline{B}}\left\{\begin{matrix} A_2 \\ A_3 \end{matrix}\right\}C \\ \underline{\underline{B}}C\left\{\begin{matrix} A_2 \\ A_3 \end{matrix}\right\} \end{matrix}\right\}$ There is no $T\ /P^{\underline{B}}/$ if $B \longrightarrow V^{ex}$

$$T\ /P^{\underline{\underline{C}}}/:\ ABC \longrightarrow \left\{\begin{matrix} A\underline{\underline{C}}B \\ \underline{\underline{C}}BA \end{matrix}\right\}$$

$$T\ /P^{\underline{C}}/:\ ABC \longrightarrow \left\{\begin{matrix} AB\underline{C} \\ BA\underline{C} \end{matrix}\right\}$$

5.1.2. The accentuation of the interrogative sentences coincides with that of the declarative ones except for one case, that is, in two-element sentences A can have secondary accent in the member BA if it is uttered with a questioning accent expressing astonishment or dubitation.

| | | | Accent |
|---|---|---|---|
| E.g. | BA: | Megy a vonat ?! | A |
| | | Megy egy vonat ?! | A |
| | | Megy vonat ?! | A |

Thus we must have some complementary rules:

$$T_Q\ /P^{\underline{A}}/:\ A_iB \longrightarrow B\underline{A_i}\ ?$$

5.2. As to the negative sentences we analyze first the declarative sentences and thereafter the interrogative sentences.

5.2.1. Of the declarative sentences sentence /1/ is first subjected to a negation transformation: $T_{nB}$, $T_{nA}$ and $T_{nAnB}$ . This is shown by the following table where x denotes the contrastive accent. The accent is designated by ___ is primary but it is not a stress accent.

/1/ {a / egy / Ø} vonat megy.

$T_{nB}$: AB → AnB

| AnB | nBA | Accent AnB | Accent nBA |
|---|---|---|---|
| A vonat nem/x/megy. | Nem/x/megy a vonat. | A B | B |
| Egy vonat nem/x/megy. | Nem/x/megy egy vonat. | A B | B |
| Vonat nem/x/megy. | Nem/x/megy vonat. | A B | B |

$T_{nA}$: AB nAB

| nAB | BnA | nAB | BnA |
|---|---|---|---|
| Nem a/x/vonat megy. | Megy nem a vonat. | A | x |
| Nem egy /x/vonat megy. | Megy nem egy vonat. | A | x |
| Nem/x/vonat megy. | Megy nem vonat. | A | x |

$T_{nAnB}$ : AB → nAnB

| nAnB | nBnA | nAnB | nBnA |
|---|---|---|---|
| Nem a/x/vonat nem megy. | Nem megy nem a vonat. | A | x |
| Nem egy/x/vonat nem megy. | Nem megy nem egy vonat. | A | x |
| Nem/x/vonat nem megy. | Nem megy nem vonat. | A | x |

We have not enumerated the permutations of /2/ because these are identical with those of /1/ except that the variant with Ø article is always incorrect. From the sentences with three elements only /6/ is analyzed:

/6/ Péter olvas/sa/ $\left\{\begin{matrix} a \\ egy \\ \emptyset \end{matrix}\right\}$ levelet.

$T_{nB}$: ABC → AnBC,

ACnB — Accent

Péter a levelet nem /x/ olvassa. C B

Péter egy levelet nem olvas. *

Péter levelet nem /x/ olvas. C B

AnBC

Péter nem /x/ olvassa a levelet. A B C

Péter nem olvas egy levelet. *

Péter nem /x/ olvas levelet. A B C

CnBA

A levelet nem /x/ olvassa Péter. C B A

Egy levelet nem olvas Péter. *

Levelet nem /x/ olvas Péter. C B A

CAnB

A levelet Péter nem /x/ olvassa. A B

Egy levelet Péter nem olvas. *

Levelet Péter nem /x/ olvas. A B

nBAC

Nem /x/ olvassa Péter a levelet. B C

Nem olvas Péter egy levelet. *

Nem /x/ olvas Péter levelet. B C

nBCA

Nem /x/ olvassa a levelet Péter. B A

Nem olvas egy levelet Péter. *

Nem /x/ olvas levelet Péter. B C

$T_{nA}$: ABC → nABC

nABC — Accent

Nem /x/ Péter olvassa a levelet. C

Nem Péter olvas egy levelet. *

Nem /x/ Péter olvas levelet. C

CnAB

A levelet nem /x/ Péter olvassa. C

Egy levelet nem Péter olvas. *

Levelet nem /x/ Péter olvas. C

$T_{nC}$: ABC → ABnC

AnCB

Péter nem /x/ a levelet olvassa. C

Péter nem egy levelet olvas. *

Péter nem /x/ levelet olvas. C

nCBA

Nem a /x/ levelet olvassa Péter. C

Nem egy levelet olvas Péter. *

Nem /x/ levelet olvas Péter. C

On the basis of the permutations we can draw the following conclusions. In the sentences with two elements in the case of $T_{nB}$ both elements of the member AnB may be accentuated, but in the member nBA only B may be accentuated, B has a negation stress accent, A, however, primary accent. If B is accentuated, then the accent falls on the negative particle, the verb may only be given contrastive accent. In the case of $T_{nA}$ and $T_{nAnB}$ the members nAB and nAnB are correct and only A can be accentuated: the negative particle obtains P, the substantive $P_{contr}$.

In the sentences with three elements B has negation stress accent in each permutation, and if the accent falls on the negative particle no complementary sentence is needed; the verb may only have contrastive accent which, however, requires a complementary sentence. The member A is given primary accent but not stress accent in the members AnBC and CAnB and secondary accent in the members CnBA and nBCA. C has primary accent but not stress accent in the members ACnB and CnBA and secondary accent in the members AnBC and nBAC.

In the sentences with three elements in the case of $T_{nA}$ only the negative particle of A may be accentuated in the members nABC and CnAB which are, incidentally, the only correct ones. In the same sentence the substantive has $P_{contr}$. After $T_{nC}$ the negative particle of C acquires P while the substantive does $P_{contr}$ in the members AnCB and nCBA.

We have not mentioned the cases of double negation /$T_{nAnB}$ , $T_{nCnB}$/ which, however, correspond to $T_{nA}$ and $T_{nC}$.

In the sentences with three elements the variant with the article <u>egy</u> is incorrect as far as the norm is concerned.

On the basis of the foregoing we can state the following rules. $\underline{\underline{\quad\quad}}$ designates the negation stress accent, $\underset{\cdot\cdot}{\underline{\quad}}$ the primary accent, $\underline{\quad}$ the secondary accent in the case of P, in the case of $P_{contr}$ x is added.

Sentences with two elements:

$T_{nB}$: AB $\rightarrow$ AnB

$$T_{nB}\ /P^{\underline{\underline{B}}}/:\ A_i nB \rightarrow \left\{\begin{matrix} A_i\underline{\underline{n}}B \\ \underline{\underline{n}}BA_i \end{matrix}\right\} \qquad T_{nB}\ /P^{\underline{\underline{B}}}_{contr}/:\ A_i nB \rightarrow \left\{\begin{matrix} A_i\underset{x}{\underline{\underline{n}}}B \\ \underset{x}{\underline{\underline{n}}}BA_i \end{matrix}\right\}$$

$$T_{nB}\ /P^{\underset{\cdot\cdot}{\underline{A}}}/:\ A_i nB \rightarrow \underset{\cdot\cdot}{\underline{A}}_i nB$$

$T_{nA}$: AB $\to$ nAB

$T_{nA}$ /$P^{\underline{\underline{A}}}$/: $nA_iB \to \underline{\underline{n}}A_iB$ $\qquad$ $T_{nA}$ $P^{\underline{\underline{A}}}_{contr}$/: $nA_iB \to \underset{x=i}{nA}B$

$T_{nAnB}$: AB $\to$ nAnB

$T_{nAnB}$ /$P^{\underline{\underline{A}}}$/: $nA_inB \to \underline{\underline{n}}A_inB$ $\qquad$ $T_{nAnB}$ /$P^{\underline{\underline{A}}}_{contr}$/: $nA_inB \to \underset{x=i}{nA}nB$

for /2/ we have the same rules but $A_i \to \begin{Bmatrix} A_1 \\ A_2 \end{Bmatrix}$

Sentences with three elements

$T_{nB}$: ABC $\to$ AnBC

$$T_{nB}/P^{\underline{\underline{B}}}/: AnBC_j \to \left\{ \begin{array}{l} A \begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix} \underline{\underline{n}}B \\ A\underline{\underline{n}}B \begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix} \\ \begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix} \underline{\underline{n}}BA \\ \begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix} A\underline{\underline{n}}B \\ \underline{\underline{n}}BA \begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix} \\ \underline{\underline{n}}B \begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix} A \end{array} \right\} \qquad TnB/P^{\underline{\underline{B}}}_{contr}/: AnBC_j \to \left\{ \begin{array}{l} A \begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix} \underset{x=}{nB} \\ A\underset{x=}{nB} \begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix} \\ \begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix} \underset{x=}{nB}A \\ \begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix} A\underset{x=}{nB} \\ \underset{x=}{nB}A \begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix} \\ \underset{x=}{nB} \begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix} A \end{array} \right\}$$

$$T_{nB}/P^{\underset{\because}{A}}/ : AnBC_j \to \left\{ \begin{array}{l} \underset{\because}{A}nB \begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix} \\ \begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix} \underset{\because}{A}nB \end{array} \right\}$$

$$T_{nB}/P^{\underline{A}}/: AnBC_j \to \left\{ \begin{array}{l} \begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix} nB\underline{A} \\ nB \begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix} \underline{A} \end{array} \right\}$$

$$T_{nB}/P^{\underline{\underline{C}}}/: AnBC_j \longrightarrow \left\{ \begin{array}{l} A\left\{\begin{array}{l}\underline{\underline{C}}_1\\ \underline{\underline{C}}_3\end{array}\right\}nB \\ \left\{\begin{array}{l}\underline{\underline{C}}_1\\ \underline{\underline{C}}_3\end{array}\right\}nBA \end{array}\right\}$$

$$T_{nB}/P^{\underline{C}}/: AnBC_j \longrightarrow \left\{ \begin{array}{l} AnB\left\{\begin{array}{l}\underline{C}_1\\ \underline{C}_3\end{array}\right\} \\ nBA\left\{\begin{array}{l}\underline{C}_1\\ \underline{C}_3\end{array}\right\} \end{array}\right\}$$

TnA: ABC $\longrightarrow$ nABC

$$T_{nA}/P^{\underline{\underline{A}}}/: nABC_j \longrightarrow \left\{ \begin{array}{l} \underline{\underline{n}}AB\left\{\begin{array}{l}C_1\\ C_3\end{array}\right\} \\ \left\{\begin{array}{l}C_1\\ C_3\end{array}\right\}\underline{\underline{n}}AB \end{array}\right\}$$

$$T_{nA}/P^{\underline{\underline{A}}}_{contr}/: nABCj \longrightarrow \left\{ \begin{array}{l} \underset{x}{n}\underline{\underline{A}}B\left\{\begin{array}{l}C_1\\ C_3\end{array}\right\} \\ \left\{\begin{array}{l}C_1\\ C_3\end{array}\right\}\underset{x}{n}\underline{\underline{A}}B \end{array}\right\}$$

$T_{nC}$: ABC $\longrightarrow$ ABnC

$$T_{nC}/P^{\underline{\underline{C}}}/: ABnC \longrightarrow \left\{ \begin{array}{l} A\underline{\underline{n}}\left\{\begin{array}{l}C_1\\ C_3\end{array}\right\}B \\ \underline{\underline{n}}\left\{\begin{array}{l}C_1\\ C_3\end{array}\right\}BA \end{array}\right\}$$

$$T_{nC}/P^{\underline{\underline{C}}}_{contr}/ \longrightarrow \left\{ \begin{array}{l} A\underset{x}{n}\left\{\begin{array}{l}\underline{\underline{C}}_1\\ \underline{\underline{C}}_3\end{array}\right\}B \\ \underset{x}{n}\left\{\begin{array}{l}\underline{\underline{C}}_1\\ \underline{\underline{C}}_3\end{array}\right\}BA \end{array}\right\}$$

5.22. The permutations of the interrogative sentences coincide with those of the declarative sentences except that in the member nBA even A may be accentuated in the case of questioning intonation expressing surprise or astonishment : Nem megy $\left\{\begin{array}{l}\underline{\underline{a}}\\ \underline{egy}\\ \emptyset\end{array}\right\}$ vonat ?

This requires a complementary rule:

$$T_{nB}/P^{\underline{\underline{A}}}/: A_i nB \longrightarrow nB\underline{A}_i$$

We leave out of account the negative questions with three elements.

5.3. Finally, we shall deal with the accentuation of the elliptical sentences.

5.3.1. We consider only one type of the elliptical declarative sentences, namely the one that can be formed from /6/ by ellipsis /6 ell/.

| BC | CB | Accent BC | Accent CB |
|---|---|---|---|
| Olvassa a levelet. | A levelet olvassa. | B C | C B |
| Olvas egy levelet. | Egy levelet olvas. | B C | C B |
| Olvas levelet. | Levelet olvas. | B C | C B |

In the member BC B may get primary accent, in the member CB the same holds true for C. B may have secondary accent in CB, and C in BC. In sentences with the verb van B cannot receive any accent /4 ell'/: $^{x}$Vagyunk a szobában. or $^{x}$ A szobában vagyunk. are incorrect sentences.

On the basis of the foregoing we can state the following rules:

$$T^{Ell/A/}\ /P^{\underline{\underline{B}}}/:\ BCj\ /I/\rightarrow \underline{\underline{B}}C_j$$

$$T^{Ell/A/}\ /P^{\underline{B}}/:\ BCj \rightarrow Cj\underline{B}$$

$$T^{Ell/A/}\ /P^{\underline{\underline{C}}}/:\ BCj \rightarrow \underline{\underline{C}}jB$$

$$T^{Ell/A/}\ /P^{\underline{C}}/:\ BCj \rightarrow B\underline{C}j$$

if $B \rightarrow V^{ex}$ , then the first two rules are out of question.

5.3.2. The permutations of the interrogative sentences coincide with those of the declarative ones therefore we do not enumerate them here. The rules that might be obtained on the basis of them are the same as those for the declarative sentences.

5.3.3. The analysis of the negative sentences is exemplified by /6 ell/.

$$T^{Ell/A/}_{nB}\ :\ BC \rightarrow nBC$$

| nBC | CnB | Accent nBC | Accent CnB |
|---|---|---|---|
| Nem$_{/x/}$olvassa a levelet. | A levelet nem$_{/x/}$olvassa. | B | C B |
| Nem olvas egy levelet. | Egy levelet nem olvas. | x | x |
| Nem$_{/x/}$olvas levelet. | Levelet nem$_{/x/}$olvas. | B | C B |

$T^{Ell/A/}_{nC}$ : BC $\to$ BnC

| BnC | nCB | Accent BnC | Accent nCB |
|---|---|---|---|
| Olvassa nem a levelet. | Nem a$_{/x/}$levelet olvassa. | * | $\underline{\underline{C}}$ |
| Olvas nem egy levelet. | Nem egy levelet olvas. | * | * |
| Olvas nem levelet. | Nem$_{/x/}$levelet olvas. | * | $\underline{\underline{C}}$ |

$T^{Ell/A/}_{nCnB}$ : BC $\to$ nBnC

| nBnC | nCnB | Accent nBnC | Accent nCnB |
|---|---|---|---|
| Nem olvassa nem a levelet. | Nem a$_{/x/}$levelet nem olvassa. | * | $\underline{\underline{C}}$ |
| Nem olvas nem egy levelet. | Nem egy levelet nem olvas. | * | * |
| Nem olvas nem levelet. | Nem$_{/x/}$levelet nem olvas. | * | $\underline{\underline{C}}$ |

In the negative sentences when B is negated then in nBC the negative particle of B, in CnB either C or the negative particle of B acquires the accent, and the contrastive accent falls in either case on the verb.

If C is negated or in the case of double negation the members BnC and nBnC are incorrect, in the member nCB the accent falls on the negative particle of C while $P_{contr}$ on the substantive, in nCnB, we have $P_{contr}$ falling on the substantive. The variants with the article egy are incorrect according to the norm. Thus we obtain the following rules:

$$T^{Ell/A/}: BC \to nBC$$

$$T^{Ell/A/}_{nB}/P^{\underline{\underline{B}}}/: nBCj \to \left\{\begin{matrix} \underline{\underline{n}}B\begin{Bmatrix}C_1\\ C_3\end{Bmatrix} \\ \\ \begin{Bmatrix}C_1\\ C_3\end{Bmatrix}\underline{\underline{n}}B \end{matrix}\right\} \qquad T^{Ell/A/}/P^{\underline{\underline{B}}}_{contr}/: nBCj \left\{\begin{matrix} \underset{x}{n}\underline{\underline{B}}\begin{Bmatrix}C_1\\ C_3\end{Bmatrix} \\ \begin{Bmatrix}C_1\\ C_3\end{Bmatrix}\underset{x}{n}\underline{\underline{B}} \end{matrix}\right\}$$

$$T^{Ell/A/}/P^{\underline{\underline{C}}}/: nBCj \to \begin{Bmatrix}\underline{\underline{C}}_1\\ \underline{\underline{C}}_3\end{Bmatrix} nB$$

$$T^{Ell/A/}_{nC}: BC \to BnC$$

$$T^{Ell/A/}_{nC}/P^{\underline{\underline{C}}}/ : BnCj \rightarrow \underset{*}{n}\begin{Bmatrix}C_1\\ C_3\end{Bmatrix} B \qquad T^{Ell/A/}_{nC}\ /P^{\underline{\underline{C}}}_{contr}/ : BnCj \rightarrow \underset{x}{\underline{\underline{n}}}\begin{Bmatrix}\underline{\underline{C}}_1\\ \underline{\underline{C}}_3\end{Bmatrix} B$$

$$T^{Ell/A/}_{nCnB} : BC \rightarrow nBnC$$

$$T^{Ell/A/}_{nCnB}\ /P^{\underline{\underline{C}}}_{contr}\ / : nBnCj \rightarrow \underset{x}{\underline{\underline{n}}}\begin{Bmatrix}\underline{\underline{C}}_1\\ \underline{\underline{C}}_3\end{Bmatrix} nB$$

5.3.4. We do not deal with the analysis of the negative interrogative sentences and with their rules because these are the same as in the case of negative sentences.

## 6. Summary

6.1. Coming to the end of our discussions of stress and accent let us summarize the rules we have seen to emerge and to state their common properties. The examination has been carried out on two levels between which a correspondence relation may be defined. On each level the affirmative and negative sentences have been considered and then - within each group - the declarative and interrogative sentences. Beside the simple two and three element sentences **also** elliptical sentences and sentences with more than three elements have been tackled. The latter, however, will be left out of account in the following. In our summary we enlist the rules of the two levels side by side.

| Stress: T/E/ | Accentuation: T/P/ |
| --- | --- |

Simple sentences

Affirmative declarative sentences /2.1.1. and 5.1.1./

Sentences with two elements

$$T/E^{o}/ : A_iB\ /I/ \rightarrow \begin{Bmatrix}A_1\\ A_2\end{Bmatrix} B \qquad T\ /P^{\underline{\underline{A}}}/ : A_iB \rightarrow \underline{\underline{A}}_iB$$

$$T/E^{\hat{A}}_{\not{o}}/ : A_iB\ /I/ \rightarrow \hat{A}_3B$$

$$T/E^{\bar{B}}/ : A_iB \rightarrow \bar{B}A_i \qquad T\ /P^{\underline{\underline{B}}}/ : A_iB \rightarrow \begin{Bmatrix}\begin{Bmatrix}A_1\\ A_2\end{Bmatrix}\underline{\underline{B}}\\ \underline{\underline{B}}A_i\end{Bmatrix}$$

Sentences with three elements

$$T\ /E^{\tilde{A}}/ : A_iBCj \rightarrow \begin{Bmatrix}\tilde{A}_1\\ \tilde{A}_2\end{Bmatrix} BCj \qquad T\ /P^{\underline{\underline{A}}}/ : ABC \rightarrow \begin{Bmatrix}\underline{\underline{A}}BC\\ C\underline{\underline{A}}B\end{Bmatrix}$$

$$T\ /E^{\bar{A}}/:\ A_iBCj \longrightarrow Cj\left\{\begin{matrix}\bar{A}_1\\ \\ \bar{A}_2\end{matrix}\right\}B$$

$$T/E^{\hat{A}}_{\emptyset}/:\ A_iBCj \longrightarrow \left\{\begin{matrix}\hat{A}_3B\left\{\begin{matrix}C_1\\ C_2\end{matrix}\right\}\\ \left\{\begin{matrix}C_1\\ C_2\end{matrix}\right\}\hat{A}_3B\end{matrix}\right\}$$

$$T\ /E^{\tilde{C}}/:\ A_iBCj \longrightarrow A_i\left\{\begin{matrix}\tilde{C}_1\\ \\ \tilde{C}_2\end{matrix}\right\}B$$

$$T\ /E^{\bar{C}}/:\ A_iBCj \longrightarrow \left\{\begin{matrix}\bar{C}_1\\ \\ \bar{C}_2\end{matrix}\right\}BA_i$$

$$T\ /E^{\hat{C}}_{\emptyset}/:\ A_iBCj \longrightarrow \left\{\begin{matrix}\left\{\begin{matrix}A_1\\ A_2\end{matrix}\right\}\hat{C}_3B\\ \hat{C}_3B\left\{\begin{matrix}A_1\\ A_2\end{matrix}\right\}\end{matrix}\right\}$$

$$T\ /E^{\bar{B}}/:\ A_iBCj \longrightarrow \left\{\begin{matrix}\bar{B}A_1Cj\\ \bar{B}CjA_1\end{matrix}\right\}$$

$$T\ /P^{\underline{A}}/:\ ABC \longrightarrow \left\{\begin{matrix}CB\underline{A}\\ \\ BC\underline{A}\end{matrix}\right\}$$

$$T\ /P^{\underline{\underline{C}}}/:\ ABC \longrightarrow \left\{\begin{matrix}A\underline{\underline{C}}B\\ \\ \underline{\underline{C}}BA\end{matrix}\right\}$$

$$T\ /P^{\underline{C}}/:\ ABC \longrightarrow \left\{\begin{matrix}AB\underline{C}\\ \\ BA\underline{C}\end{matrix}\right\}$$

$$T\ /P^{\underline{\underline{B}}}/:\ ABC \longrightarrow \left\{\begin{matrix}\underline{\underline{B}}AC\\ \\ \underline{\underline{B}}CA\end{matrix}\right\}$$

$$T\ /P^{\underline{B}}/:\ ABC \longrightarrow \left\{\begin{matrix}ACB\\ AB\bar{C}\\ C\bar{B}A\\ C\bar{A}\underline{B}\end{matrix}\right\}$$

Negative declarative sentences /2.2.1. and 5.2.1./

Sentences with two elements

$$T_{nB}\ :\ AB \longrightarrow AnB$$

$$T_{nB}\ /E^{\tilde{B}}/:\ A_inB\ /I/ \longrightarrow A_in\tilde{B}$$

$$T_{nB}\ /P^{\underline{\underline{B}}}/:\ A_inB \longrightarrow \left\{\begin{matrix}A_i\underline{\underline{n}}B\\ \\ \underline{\underline{n}}BA_i\end{matrix}\right\}$$

$$T_{nB}\ /E^{\bar{B}}/:\ A_inB \longrightarrow n\bar{B}A_i$$

$$T_{nB}\ /P^{\underline{\underline{B}}}_{contr}/:\ A_inB \longrightarrow \left\{\begin{matrix}A_i\underset{x}{\underline{\underline{nB}}}\\ \\ \underset{x}{\underline{\underline{nB}}}A_i\end{matrix}\right\}$$

$$T_{nB}\ /P^{\underset{..}{\underline{\underline{A}}}}/:\ A_i nB \rightarrow \underset{..}{\underline{\underline{A}}}_i nB$$

$$T_{nA}\ :\ AB \rightarrow nAB$$

$$T_{nA}\ /E^{\tilde{A}}/:\ nA_iB \rightarrow n\begin{Bmatrix}\tilde{A}_1\\ \tilde{A}_2\end{Bmatrix} B \qquad\qquad T_{nA}\ /P^{\underline{\underline{A}}}/:\ nA_iB \rightarrow \underline{\underline{n}}A_1B$$

$$T_{nA}\ /E^{\hat{A}}_{\emptyset}/:\ nA_iB \rightarrow n\hat{A}_3B \qquad\qquad T_{nA}\ /P^{\underline{\underline{A}}}_{contr}/:\ nA_iB \rightarrow \underset{x}{\cancel{n}}\underline{\underline{A}}_iB$$

$$T_{nAnB}:\ AB \rightarrow nAnB$$

$$T_{nAnB}\ /E^{\tilde{A}}/:\ nA_inB \rightarrow n\begin{Bmatrix}\tilde{A}_1\\ \tilde{A}_2\end{Bmatrix} nB \qquad\qquad T_{nAnB}/P^{\underline{\underline{A}}}/:\ nA_inB \rightarrow \underline{\underline{n}}A_inB$$

$$T_{nAnB}\ /E^{\hat{A}}_{\emptyset}/:\ nA_inB \rightarrow n\hat{A}_3nB \qquad\qquad T_{nAnB}\ /P^{\underline{\underline{A}}}_{contr}/:\ nA_inB \rightarrow \underset{x}{\cancel{n}}\underline{\underline{A}}_inB$$

Sentences with three elements

$$T_{nB}:\ ABC \rightarrow AnBC$$

$$T_{nB}/P^{\tilde{\bar{B}}}/:\ AnBC_j \rightarrow \begin{Bmatrix} A\begin{Bmatrix}C_1\\ C_3\end{Bmatrix}n\tilde{B} \\ An\tilde{B}\begin{Bmatrix}C_1\\ C_3\end{Bmatrix} \\ \begin{Bmatrix}C_1\\ C_3\end{Bmatrix}n\tilde{B}A \\ \begin{Bmatrix}C_1\\ C_3\end{Bmatrix}An\tilde{B} \end{Bmatrix}$$

$$T_{nB}/P^{\underline{\underline{B}}}/:\ AnBC_j \rightarrow \begin{Bmatrix} A\begin{Bmatrix}C_1\\ C_3\end{Bmatrix}\underline{\underline{n}}B \\ A\underline{\underline{n}}B\begin{Bmatrix}C_1\\ C_3\end{Bmatrix} \\ \begin{Bmatrix}C_1\\ C_3\end{Bmatrix}\underline{\underline{n}}BA \\ \begin{Bmatrix}C_1\\ C_3\end{Bmatrix}A\underline{\underline{n}}B \\ \underline{\underline{n}}BA\begin{Bmatrix}C_1\\ C_3\end{Bmatrix} \\ \underline{\underline{n}}B\begin{Bmatrix}C_1\\ C_3\end{Bmatrix}A \end{Bmatrix}$$

$$T_{nB}/P^{\bar{\bar{B}}}/:\ AnBC_j \longrightarrow \left\{\begin{array}{l} n\bar{\bar{B}}A\begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix} \\[2ex] n\bar{\bar{B}}\begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix}A \end{array}\right\}$$

$$T_{nB}/P^{\underline{\underline{B}}}_{contr}/:\ AnBC_j \longrightarrow \left\{\begin{array}{l} A\begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix}\underset{x}{n}\underline{\underline{B}} \\[1ex] A\underset{x}{n}\underline{\underline{B}}\begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix} \\[1ex] \begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix}\underset{x}{n}\underline{\underline{B}}A \\[1ex] \begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix}A\underset{x}{n}\underline{\underline{B}} \\[1ex] \underset{x}{n}\underline{\underline{B}}A\begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix} \\[1ex] \underset{x}{n}\underline{\underline{B}}\begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix} \end{array}\right\}$$

$$T_{nB}/P\underset{\cdot\cdot}{\underline{\underline{A}}}/:\ AnBC_j \longrightarrow \left\{\begin{array}{l} \underset{\cdot\cdot}{\underline{\underline{A}}}nB\begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix} \\[2ex] \begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix}\underset{\cdot\cdot}{\underline{\underline{A}}}nB \end{array}\right\}$$

$$T_{nB}/P\underline{\underline{A}}/:\ AnBC_j \longrightarrow \left\{\begin{array}{l} \begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix}nB\underline{\underline{A}} \\[2ex] nB\begin{Bmatrix} C_1 \\ C_3 \end{Bmatrix}\underline{\underline{A}} \end{array}\right\}$$

$$T_{nB}/P^{\underset{\cdots}{\underline{C}}}/: AnBC_j \rightarrow \left\{ \begin{array}{l} A \left\{\begin{array}{l} \underset{\cdots}{\underline{C}}_1 \\ \underset{\cdots}{\underline{C}}_3 \end{array}\right\} nB \\ \left\{\begin{array}{l} \underset{\cdots}{\underline{C}}_1 \\ \underset{\cdots}{\underline{C}}_3 \end{array}\right\} nBA \end{array}\right\}$$

$$T_{nB}/P^{\underline{C}}/: AnBC_j \rightarrow \left\{ \begin{array}{l} AnB \left\{\begin{array}{l} \underline{C}_1 \\ \underline{C}_3 \end{array}\right\} \\ nBA \left\{\begin{array}{l} \underline{C}_1 \\ \underline{C}_3 \end{array}\right\} \end{array}\right\}$$

$T_{nA}$: ABC $\rightarrow$ nABC

$$T_{nA}/E^{\tilde{A}}/: nABC_j \rightarrow n\tilde{A}B \left\{\begin{array}{l} C_1 \\ C_3 \end{array}\right\}$$

$$T_{nA}/E^{\bar{A}}/: nABC_j \rightarrow \left\{\begin{array}{l} C_1 \\ C_3 \end{array}\right\} n\bar{A}B$$

$$T_{nA}/P^{\underline{\underline{A}}}/: nABC_j \rightarrow \left\{ \begin{array}{l} \underline{\underline{n}}AB \left\{\begin{array}{l} C_1 \\ C_3 \end{array}\right\} \\ \left\{\begin{array}{l} C_1 \\ C_3 \end{array}\right\} \underline{\underline{n}}AB \end{array}\right\}$$

$$T_{nA}/P^{\underline{\underline{A}}}_{contr}/: nABC_j \rightarrow \left\{ \begin{array}{l} \underset{x}{\underline{\underline{n}}}AB \left\{\begin{array}{l} C_1 \\ C_3 \end{array}\right\} \\ \left\{\begin{array}{l} C_1 \\ C_3 \end{array}\right\} \underset{x}{\underline{\underline{n}}}AB \end{array}\right\}$$

$T_{nC}$: ABC $\rightarrow$ ABnC

$$T_{nC}/E^{\tilde{C}}/: ABnC_j \rightarrow An\tilde{C}_1B$$

$$T_{nC}/E^{\bar{C}}/: ABnC_j \rightarrow n\bar{C}_1BA$$

$$T_{nC}/E^{\hat{C}}_{\emptyset}/: ABnC_j \rightarrow \left\{\begin{array}{l} An\hat{C}_3B \\ n\hat{C}_3BA \end{array}\right\}$$

$$T_{nC}/P^{\underline{\underline{C}}}/: ABnC \rightarrow \left\{ \begin{array}{l} A\underline{\underline{n}} \left\{\begin{array}{l} C_1 \\ C_3 \end{array}\right\} B \\ \underline{\underline{n}} \left\{\begin{array}{l} C_1 \\ C_3 \end{array}\right\} BA \end{array}\right\}$$

$$T_{nC}/P^{\underline{\underline{C}}}_{contr}/ \rightarrow \left\{ \begin{array}{l} A\underset{x}{\underline{\underline{n}}} \left\{\begin{array}{l} \underline{\underline{C}}_1 \\ \underline{\underline{C}}_3 \end{array}\right\} B \\ \underset{x}{\underline{\underline{n}}} \left\{\begin{array}{l} \underline{\underline{C}}_1 \\ \underline{\underline{C}}_3 \end{array}\right\} BA \end{array}\right\}$$

$T_{nAnB}$: ABC $\rightarrow$ nAnBC see $T_{nA}$

$T_{nCnB}$: ABC $\rightarrow$ AnBnC see $T_{nC}$

## Elliptical sentences

### Affirmative declarative sentences /3.1. and 5.3.1./

$$T^{Ell/A/}: ABC \rightarrow BC$$

$T^{Ell/A/}\ /E^{\tilde{B}}/:\ BC_j\ /I/ \rightarrow \tilde{B}Cj$

$T^{Ell/A/}\ /E^{\tilde{C}}/:\ BCj \rightarrow \left\{\begin{matrix}\tilde{C}_1\\ \tilde{C}_2\end{matrix}\right\}B$

$T^{Ell/A/}\ /E_{\emptyset}^{\hat{C}}/:\ BCj \rightarrow \hat{C}_3B$

$T^{Ell/A/}\ /P^{\underline{\underline{B}}}/:\ BCj\ /I/ \rightarrow \underline{\underline{B}}Cj$

$T^{Ell/A/}\ /P^{\underline{B}}/:\ BCj \rightarrow Cj\underline{\underline{B}}$

$T^{Ell/A/}\ /P^{\underline{\underline{C}}}/:\ BCj \rightarrow \underline{\underline{C}}jB$

$T^{Ell/A/}\ /P^{\underline{C}}/:\ BCj \rightarrow B\underline{\underline{C}}j$

### Negative declarative sentences /3.2 and 5.3.3./

$$T_{nB}^{Ell/A/}: BC \rightarrow nBC$$

$T_{nB}^{Ell/A/}\ /E^{\tilde{B}}/:\ nBCj \rightarrow \left\{\begin{matrix}C_1\\ C_3\end{matrix}\right\}n\tilde{B}$

$T_{nB}^{Ell/A/}\ /E^{\bar{B}}/:\ nBCj — n\bar{B}\left\{\begin{matrix}C_1\\ C_3\end{matrix}\right\}$

$T_{nB}^{Ell/A/}/P^{\underline{\underline{B}}}/:\ nBCj \rightarrow \left\{\begin{matrix}\underline{\underline{n}}B\left\{\begin{matrix}C_1\\ C_3\end{matrix}\right\}\\ \left\{\begin{matrix}C_1\\ C_3\end{matrix}\right\}\underline{\underline{n}}B\end{matrix}\right\}$

$T_{nB}^{Ell/A/}/P_{contr}^{\underline{\underline{B}}}/:\ nBCj \rightarrow \left\{\begin{matrix}\underset{x}{\underline{\underline{nB}}}\left\{\begin{matrix}C_1\\ C_3\end{matrix}\right\}\\ \left\{\begin{matrix}C_1\\ C_3\end{matrix}\right\}\underset{x}{\underline{\underline{nB}}}\end{matrix}\right\}$

$T_{nB}^{Ell/A/}\ /P^{\underline{C}}/:\ nBCj \rightarrow \left\{\begin{matrix}\underline{C}_1\\ \underline{C}_3\end{matrix}\right\}nB$

$$T_{nC}^{Ell/A/}: BC \rightarrow BnC$$

$T_{nC}^{Ell/A/}\ /E^{\tilde{C}}/:\ BnCj \rightarrow n\tilde{C}_1B$

$T_{nC}^{Ell/A/}\ /E_{\alpha}^{\hat{C}}/:\ BnCj \rightarrow n\hat{C}_3B$

$T_{nC}^{Ell/A/}/P^{\underline{\underline{C}}}/:\ BnCj \rightarrow \underline{\underline{n}}\left\{\begin{matrix}C_1\\ C_3\end{matrix}\right\}B$

$T_{nC}^{Ell/A/}\ /P_{contr}^{\underline{\underline{C}}}/:\ BnCj \rightarrow \underset{x}{\underline{\underline{n}}}\left\{\begin{matrix}\underline{\underline{C}}_1\\ \underline{\underline{C}}_3\end{matrix}\right\}B$

$$T^{Ell/A/}_{nBnC}: BC \rightarrow nBnC$$

see $T^{Ell/A/}_{nC}$ see $T^{Ell/A/}_{nC}$

6.2. If we have a look at the sequences of transformations enumerated above, the question arises whether it would be possible to find comprehensive rules for the great variety of transformations because there is obviously a system behind this variety. Affirmative declarative sentences are regarded as primary, the negative declarative sentences mainly coincide with the former and depart from them only in some rules. The interrogative sentences correspond to the declarative ones only slightly departing from them. The rules of accentuation are more general, the stress rules can be derived from them. The general rule for the declarative accentuation can be given in the following form:

$$T\ /P^{\underline{\underline{X}}}/: WXY \rightarrow \begin{Bmatrix} W\underline{\underline{X}}Y \\ \underline{\underline{X}}YW \end{Bmatrix}$$

$$T\ /P^{\underline{\underline{Y}}}/: WXY \rightarrow \begin{Bmatrix} \underline{\underline{Y}}XW \\ \underline{\underline{Y}}WX \end{Bmatrix}$$

where $\begin{Bmatrix} W \\ X \end{Bmatrix} \rightarrow \begin{Bmatrix} \emptyset \\ A \\ C \end{Bmatrix}$, $Y \rightarrow B$,

Thus X is accentuated if it stands before Y and on the other hand, Y is accentuated, if it is shifted to the beginning of the sentence. The rules of the secondary accentuation:

$$T\ /P^{\underline{X}}/: WXY \rightarrow \begin{Bmatrix} WY\underline{X} \\ YW\underline{X} \end{Bmatrix}$$

$$T\ /P^{\underline{Y}}/: WXY \rightarrow \begin{Bmatrix} W\underline{Y}X \\ WX\underline{Y} \end{Bmatrix}$$

Thus X has secondary accent if it stands after Y at the end of the sentence and Y has secondary accent, if it does not stand at the beginning of the sentence, so the preceding W, X may be accentuated.

The stress falls on the same element as the primary accent in the case of X. If $X \rightarrow \begin{Bmatrix} X_1 \\ X_2 \\ X_3 \end{Bmatrix}$, $/E^{X}/$ falls on $X_1$, $X_2$, and $/E^{X}_{b}/$ on $X_3$.

In the case of negation we have the following accent rules:

$$T_{nX}\ /P^{\underline{\underline{X}}}/:\ WnXY \rightarrow \begin{Bmatrix} W\underline{\underline{n}}XY \\ \underline{\underline{n}}XYW \end{Bmatrix}$$

$$T_{nY}\ /P^{\underline{\underline{Y}}}/:\ WXnY \rightarrow \begin{Bmatrix} X\underline{\underline{n}}YW \\ X\overline{W}\underline{\underline{n}}Y \\ \underline{\underline{n}}YXW \end{Bmatrix}$$

$$T_{nY}\ /P^{\underset{\cdot\cdot}{\underline{\underline{X}}}}/:\ WXnY \rightarrow \begin{Bmatrix} W\underset{\cdot\cdot}{\underline{\underline{X}}}nY \\ \underset{\cdot\cdot}{\underline{\underline{X}}}nYW \end{Bmatrix}$$

$$T_{nY}\ /P^{\underline{X}}/:\ WXnY \rightarrow \begin{Bmatrix} WnY\underline{X} \\ nYW\underline{X} \end{Bmatrix}$$

Thus if X is negated, then the negative particle receives stress accent, if X precedes Y. If Y is negated, the negative particle of Y is given stress accent in any position. X acquires primary but not stress accent, if it precedes nY and a secondary accent, if it stands at the end of the sentence. The rules of the contrastive accentuation coincide with the first two groups of rules but the accent does not fall on the negative particle but on X or Y. Other contrastive accents may not occur:

$$T_{nX}\ /P^{\underline{\underline{X}}}_{contr}/:\ WnXY \rightarrow \begin{Bmatrix} W\xcancel{n}\underline{\underline{X}}Y \\ \xcancel{n}\underline{\underline{X}}YW \end{Bmatrix}$$

$$T_{nY}\ /P^{\underline{\underline{Y}}}_{contr}/:\ WXnY \rightarrow \begin{Bmatrix} X\xcancel{n}\underline{\underline{Y}}W \\ XW\xcancel{n}\underline{\underline{Y}} \\ \xcancel{n}\underline{\underline{Y}}XW \end{Bmatrix}$$

The place of stress coincides with that of the stress accent. The rules given above are general but not without exception. So, for instance, if $Y \rightarrow V^{ex}$ then there are only restricted possibilities of accentuation and in negative sentences the variant with the article egy will be incorrect etc.

7. The syntactic and pragmatic analysis of word order.

In the foregoing our investigations were confined to the frames of the sentence and we stated the rules of stressing and accentuation. But of the stressed sentences those in which the stress is stronger, i.e. where the stress has been formed by the rules $T\ /E^{A}/$, $T\ /E^{B}/$, $T\ /E^{C}/$ and all the accentuated sentences are organically connected with the preceding text, they carry a stage further the message contained in the foregoing text and convey new information. Therefore it is necessary to examine the actual or logical segmentation of the sentence which we have included among the pragmatic investigations. As pointed out such investigation has a long and rich tradition in Hungarian linguistic literature. The pragmatic investigations were cultivated above all by the linguists of the Prague circle. Among the scholars who have contributed to the research in generative grammar E. Bach and D.S, Worth have repeatedly pointed to the importance of this aspect of investigation and they have offered some important remarks in this respect. K. Heidolph has written a monograph /not published so far/ on the word order in German in which the pragmatic aspects are treated in detail.

We do not aim at the examination of the actual segmentation instead we content ourselves with pointing to some interconnections between the results of the present work and the analysis of actual sentences. An article by László Elekfi deals at length with the actual analysis of Hungarian sentences with two and three elements and we base our remarks on his findings. It should be made clear, however, that we approach the problems from our own point of view and Elekfi's results are formalized according to our conventions.

Elekfi divides the sentences into two groups, in one group the theme stands in the first place and the propositum in the second: A regényt olvastam. Olvastam a regényt. /Elekfi 338/, this is the rational sentence, in the other group the situation is reversed: the theme is preceded by the propositum, such sentences are emotional: A regényt olvastam. Olvastam a regényt. Thus, if X and Y denotes any member, we have

$$S^{rac} \rightarrow \begin{Bmatrix} X\underline{Y} \\ Y\underline{X} \end{Bmatrix}, \quad S^{em} \rightarrow \begin{Bmatrix} \underline{X}Y \\ \underline{Y}X \end{Bmatrix}$$

In this publication Elekfi deals only with the first type but conclusions concerning the second type, can be made on the basis of our material.

It is clear from our evidence that the complete /not defective/ sentence with two elements has the following accentuation: $A\underline{B}$, $B\underline{A}$, this is in full accord with the conclusions reached by Elekfi. We find, the following two types in his article:

$$S^{rac'} \rightarrow N_n + \begin{Bmatrix} \underline{\underline{V}} \\ \underline{\underline{N'_n}} \end{Bmatrix}, \qquad S^{rac''} \rightarrow \begin{Bmatrix} V \\ Adj \\ N'_n \end{Bmatrix} + \underline{N}_n$$

$$S^{rac'}_{ell/A/} \rightarrow \begin{Bmatrix} N_x \\ Adv \end{Bmatrix} + \underline{V}, \qquad S^{rac''}_{ell/A/} \rightarrow V + \underline{N}_x$$

Here only those rules are given that have correspondent sentences in Elekfi, the order is that of the actual sentence. Elekfi did not aim at completeness and his work can be complemented in several respects. The part-of-speech symbols can be rewritten as symbols A,B,C in the following way·

$$S^{rac'} \rightarrow A\underline{\underline{B}} \qquad S^{rac''} \rightarrow B\underline{A}$$

$$S^{rac'}_{ell/A/} \rightarrow C\underline{B} \qquad S^{rac''}_{ell/A/} \rightarrow B\underline{C}$$

These can be formed by means of the following rules:

$S^{rac'}$ by $T\ /P^{\underline{\underline{B}}}/$; $S^{rac''}$ by $T\ /P^{\underline{A}}/$; $S^{rac'}_{ell/A/}$ by $T^{Ell/A/}/P^{\underline{B}}/$;

$S^{rac''}_{ell/A/}$ by $T^{Ell/A/}\ /P^{\underline{C}}/$.

In the sentences with three elements the first element is the theme, the other two elements the propositum. If the first member of the propositum is accentuated, then the sentence is rational /emotional / , if the second obtains the accent, then it is rational /rational/.

$$S^{rac/em/} \rightarrow \begin{Bmatrix} X\underline{Y}W \\ X\underline{W}Y \end{Bmatrix}, \qquad S^{rac/rac/} \rightarrow \begin{Bmatrix} XY\underline{W} \\ XW\underline{Y} \end{Bmatrix}$$

On the basis of the examples given by Elekfi we can draw the following conclusions /we consider only the complete affirmative sentences/:

$$S^{rac/em/} \rightarrow \left\{ \begin{array}{lll} N_n & \begin{Bmatrix} \underline{\underline{N}}_x \\ Adv \\ \underline{\underline{Adj}} \\ \underline{\underline{N'_n}} \end{Bmatrix} & V \\ N_n & \underline{V} & N_a \\ & \begin{Bmatrix} Adv \\ N_x \end{Bmatrix} & \underline{\underline{N}}_n V \end{array} \right\}$$

$$S^{rac/rac/} \rightarrow \left\{ \begin{array}{ll} N_n N_x & \begin{Bmatrix} \underline{V} \\ \underline{Adv} \\ \underline{N'_n} \end{Bmatrix} \\ \begin{Bmatrix} N_x \\ Adv \end{Bmatrix} & N_n\underline{V} \\ N_n V \underline{N}_x & \\ N_x V \underline{N}_n & \end{array} \right\}$$

If these are rewritten by the symbols A,B,C, we see that

$$S^{rac/em/} \rightarrow \left\{ \begin{matrix} A\underline{\underline{C}}B \\ A\underline{B}C \\ C\underline{\underline{A}}B \end{matrix} \right\}, \qquad S^{rac/rac/} \rightarrow \left\{ \begin{matrix} AC\underline{B} \\ CA\underline{B} \\ AB\underline{C} \\ CB\underline{A} \end{matrix} \right\}$$

These may be formed with the help of the following rules

$$S^{rac/em/}: \begin{matrix} A\underline{\underline{C}}B \text{ by } T\ /P^{\underline{\underline{C}}}/ \\ A\underline{B}C \text{ by } T\ /P^{\underline{B}}/ \\ C\underline{\underline{A}}B \text{ by } T\ /P^{\underline{\underline{A}}}/ \end{matrix} \qquad S^{rac/rac/}: \begin{matrix} AC\underline{B},\ CA\underline{B} \text{ by } T\ /P^{\underline{B}}/ \\ AB\underline{C} \text{ by } T\ /P^{\underline{C}}/ \\ CB\underline{A} \text{ by } T\ /P^{\underline{A}}/ \end{matrix}$$

Of the $T\ /P^{\underline{B}}/$ rules, however, in the first case only one,in the second case only two could be applied, from the rules $T\ /P^{\underline{C}}/$, $T\ /P^{\underline{A}}/$ only one which can be guaranteed by indexing the rules. It is, however, not clear whether the other rules would not also be necessitated if a larger corpus is considered. It can be stated that the number of permutations which can be accentuated in the second and third places exceeds the number given by the examples.

The emotional sentences obviously do not belong to the rational sentences. So, in the case of sentences with two elements

$$S^{em} \rightarrow \left\{ \begin{matrix} \underline{\underline{A}}B \\ \underline{\underline{B}}A \end{matrix} \right\}$$

In the case of sentences with three elements one has to take into consideration that the element standing in the first place in ACB and CAB cannot be accentuated and correspondingly the following sentence types and rules are necessary:

$$S^{em} \rightarrow \left\{ \begin{matrix} \underline{\underline{A}}BC \\ \underline{\underline{C}}BA \\ \underline{\underline{B}}AC \\ \underline{\underline{B}}CA \end{matrix} \right\} \qquad S^{em}: \begin{matrix} T\ /P^{\underline{\underline{A}}}/ \\ T\ /P^{\underline{\underline{C}}}/ \\ T\ /P^{\underline{\underline{B}}}/ \end{matrix}$$

With this we conclude the analysis of the sentence types of actual segmentation which is to be considered only as a starting point for a fuller examination of this matter. The investigations in this field have to aim at stating those rules by means of which it is possible to decide on the basis of the preceding text which sentence type has to be used.

Appendix 1.

/1/ $\left\{\begin{matrix}\underline{a}\\ \underline{egy}\\ \emptyset\end{matrix}\right\}$ <u>vonat megy.</u> ' $\left\{\begin{matrix}the\\ a\\ \emptyset\end{matrix}\right\}$ train goes'.

/A $\rightarrow$ T + $N_N$, V $\rightarrow$ $V^{intr}$, n $\rightarrow$ <u>nem</u> /

| | | | | Declarative | | Question | |
|---|---|---|---|---|---|---|---|
| | AB | | BA | AB | BA | AB | BA |
| AnB: | aNnV | nBA: | nVaN | B | B | B | B |
| | eNnV | | nVeN | B | B | B | B |
| | øNnV | | nVøN | B | B | B | B |
| nAB: | naNV | BnA: | VnaN | A | ⁎ | A | ⁎ |
| | neNV | | VneN | A | ⁎ | A | ⁎ |
| | nøNV | | VnøN | A | ⁎ | A | ⁎ |
| nAnB: | naNnV | nBnA: | nVnaN | A | ⁎ | A | ⁎ |
| | neNnV | | nVneN | A | ⁎ | A | ⁎ |
| | nøNnV | | nVnøN | A | ⁎ | A | ⁎ |

/2/ $\left\{\begin{matrix}\underline{a}\\ \underline{egy}\\ \emptyset\end{matrix}\right\}$ <u>ház magas.</u> ' $\left\{\begin{matrix}the\\ a\\ \emptyset\end{matrix}\right\}$ house is high'.

| | | | | Declarative | | Question | |
|---|---|---|---|---|---|---|---|
| | AB | | BA | AB | BA | AB | BA |
| AnB: | aNnAdj | nBA: | nAdjaN | B | B | B | B |
| | eNnAdj | | nAdjeN | B | B | B | B |
| | øNnAdj | | nAdjøN | ⁎ | ⁎ | ⁎ | ⁎ |
| nAB: | naNAdj | BnA: | AdjnaN | A | ⁎ | A | ⁎ |
| | neNAdj | | AdjneN | A | ⁎ | A | ⁎ |
| | nøNAdj | | AdjnøN | ⁎ | ⁎ | ⁎ | ⁎ |
| nAnB: | naNnAdj | nBnA: | nAdjnaN | A | ⁎ | A | ⁎ |
| | neNnAdj | | nAdjneN | A | ⁎ | A | ⁎ |
| | nøNnAdj | | nAdjnøN | ⁎ | ⁎ | ⁎ | ⁎ |

Appendix 2.

/6/ Péter olvas/sa/ $\left\{\begin{matrix} \text{a} \\ \text{egy} \\ \emptyset \end{matrix}\right\}$ levelet. 'Péter reads $\left\{\begin{matrix} \text{the} \\ \text{a} \\ \emptyset \end{matrix}\right\}$ letter'.

/A $\rightarrow N_N^{prop}$  B $\rightarrow V^{tr}$ ,  C $\rightarrow$ T+$N_A$  n $\rightarrow$ nem/

| nB | nA | nC | nB | nA | nC |
|---|---|---|---|---|---|
| $N_N$ $aN_A$ nV | $nN_N$ $aN_A$ V | $N_N$ $naN_A$ V | B | * | C |
| $N_N$ $eN_A$ nV | $nN_N$ $eN_A$ V | $N_N$ $neN_A$ V | * | * | * germ |
| $N_N$ $\emptyset N_A$ nV | $nN_N$ $\emptyset N_A$ V | $N_N$ $n\emptyset N_A$ V | B | * | C |
| $N_N$ nV $aN_A$ | $nN_N$ V $aN_A$ | $N_N$ V $naN_A$ | B | A | * |
| $N_N$ nV $eN_A$ | $nN_N$ V $eN_A$ | $N_N$ V $neN_A$ | * | A | * |
| $N_N$ nV $\emptyset N_A$ | $nN_N$ V $\emptyset N_A$ | $N_N$ V $n\emptyset N_A$ | B | /*/A | * |
| $aN_A$ nV $N_N$ | $aN_A$ V $nN_N$ | $naN_A$ V $N_N$ | B | * | C |
| $eN_A$ nV $N_N$ | $eN_A$ V $nN_N$ | $neN_A$ V $N_N$ | * | * | * germ |
| $\emptyset N_A$ nV $N_N$ | $\emptyset N_A$ V $nN_N$ | $n\emptyset N_A$ V $N_N$ | B | * | C |
| $aN_A$ $N_N$ nV | $aN_A$ $nN_N$ V | $naN_A$ $N_N$ V | B | A | * |
| $eN_A$ $N_N$ nV | $eN_A$ $nN_N$ V | $neN_A$ $N_N$ V | * | * | * |
| $\emptyset N_A$ $N_N$ nV | $\emptyset N_A$ $nN_N$ V | $n\emptyset N_A$ $N_N$ V | B | A | * |
| nV $N_N$ $aN_A$ | V $nN_N$ aNA | V $N_N$ $naN_A$ | B | * | * |
| nV $N_N$ $eN_A$ | V $nN_N$ $eN_A$ | V $N_N$ $neN_A$ | * | * | * |
| nV $N_N$ $\emptyset N_A$ | V $nN_N$ $\emptyset N_A$ | V $N_N$ $n\emptyset N_A$ | B | * | * |
| nV $aN_A$ $N_N$ | V $aN_A$ $nN_N$ | V $naN_A$ $N_N$ | B | * | * |
| nV $eN_A$ $N_N$ | V $eN_A$ $nN_N$ | V $neN_A$ $N_N$ | * | * | * |
| nV $\emptyset N_A$ $N_N$ | V $\emptyset N_A$ $nN_N$ | V $n\emptyset N_A$ $N_N$ | B | * | * |

germ. = Germanism

Appendix 3.

/6 ell/ <u>olvas/sa/</u> $\left\{\begin{matrix}\underline{a}\\ \underline{egy}\\ \underline{\emptyset}\end{matrix}\right\}$ <u>levelet.</u>* 'Reads $\left\{\begin{matrix}\text{the}\\ \text{a}\\ \emptyset\end{matrix}\right\}$ letter'

/ $B \rightarrow V^{tr}$, $C \rightarrow T + N_A$ /

| | 1. Column | | 2. column | 1. col. | 2. col. | 1. col. | 2.c. |
|---|---|---|---|---|---|---|---|
| BC: | $VaN_A$ | CB: | $aN_AV$ | B | C | B | C |
| | $VeN_A$ | | $eN_AV$ | B | C | B | C |
| | $V\emptyset N_A$ | | $\emptyset N_AV$ | /*/ B | C | B | C |
| nBC: | $nVaN_A$ | CnB: | $aN_AnV$ | B | B | B | C |
| | $nVeN_A$ | | $eN_AnV$ | * | * | * | * |
| | $nV\emptyset N_A$ | | $\emptyset N_AnV$ | B | B | B | B |
| BnC: | $VnaN_A$ | nCB: | $naN_AV$ | * | C | * | C |
| | $VneN_A$ | | $neN_AV$ | * | * | * | * |
| | $Vn\emptyset N_A$ | | $n\emptyset N_AV$ | * | C | * | C |
| nBnC: | $nVnaN_A$ | nCnB: | $naN_AnV$ | * | C | * | C |
| | $nBneN_A$ | | $neN_AnV$ | * | * | * | * |
| | $nVn\emptyset N_A$ | | $n\emptyset N_AnV$ | * | C | * | C |

* This identical to /3 ell'/: <u>Megyünk</u> $\left\{\begin{matrix}\underline{az}\\ \underline{egy}\\ \underline{\emptyset}\end{matrix}\right\}$ <u>iskolába.</u>

/ $B \rightarrow V^{intr}$, $C \rightarrow T + N_{ba}$ /

Notes

1. The present article is a shortened version of the first part of a comprehensive monograph written in Hungarian. In this part only those sentences are considered, which are made with verbs without verbal prefix. The analysis of sentences with prefixal verbs and the typological investigation of Hungarian word order will be the subject of the next two parts of the monograph.
2. The references to the most important works by the mentioned authors are to be found in the bibliography. Further references may be found in the given works.
3. The evaluation of this important trend falls outside of the scope of the present work. The article by O.A.Lapteva: Českoslovackije raboty poslednich let po voprosam aktual'nogo členenija predloženija. Voprosy Jazykoznanija 13:4.12o-7. /1963 /. It is reviewed by L.Antal, Word Order and Syntactic Position. Linguistics 8. 31-42.
4. We are indebted to P.Engel, I. Köves and G. Pálvölgyi who took upon themselves the fatiguing labour of acting as informants. Grateful acknowledgements are due to my colleagues S.Ábrahám, L.Elekfi, S.Károly, F.Kiefer, I.Molnár, S.J.Petőfi, Gy.Szépe, Zs.Telegdi for their valuable remarks and comments.
5. Dénes Varga is engaged in work in this direction in connection with MT from Russian into Hungarian.

## Bibliography


Adamec, P., K rozdílem mezi českým a ruským slovosledem. Kapitoly ze srovnávací mluvnice ruské a české. I. Praha, 1956 . 180-198.

Arany, J., A szórend. Magyar Nyelvőr II. 7-11 /1873/.

Bach, E., The Order of Elements in a Transformational Grammar of German. Language 38. 263-9 / 1962 / .

Bernecker, E., Die Wortfolge in den slavischen Sprachen. Berlin, 1900.

Boost, K., Neue Untersuchungen zum Wesen und zur Struktur des deutschen Satzes. Berlin, 1955.

Brassai, S., Szórend és akcentus. Budapest, 1888.

Buttke, K., Besonderheiten der Wortfolge in Stilen der russischen Literatursprache der Gegenwart. Zeitschrift für Slawistik 7: 785--792 /1963/ .

Daneš, F., K otázce pořádku slov v slovanských jazycích. Slovo a slovesnost 20. 1--10 /1959/ .

-- Intonace a věta ve spisovné češtine. Praha, 1957.

Deme, L., A mai magyar nyelv rendszere. Leíró nyelvtan. II. Bp. 1962, 471--502.

Elekfi, L., Az aktuális mondattagolás egyik alapformája a magyarban. Nyelvtudományi Közlemények 66. 331-370 /1964/ .

Fogarasi, J., A szóhelyezésről a magyar nyelvben. Atheneum II. 193-198, 240-249, 289-297 /1838/ .

Gabelentz, G.v., Die Sprachwissenschaft, ihre Aufgaben, Methoden und bisherigen Ergebnisse. Leipzig, 1891.

Greenberg, J., Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements: Universals of Language. Cambridge, Mass. 1963. 58-90.

Hetzron, R. L'accent en hongrois. Bulletin de la Société de Linguistique de Paris 57. 192-205 /1962/ .

Joannovics, Gy., Szórend. Magyar Nyelvőr I. 166-71, 206-14. 254-60, 303-10 /1872/ , XIII. 53-8, 103-10, 197-204, 299-306 /1884/.

Krušelnickaja, K.G., K voprosu o smyslovom členenii predloženija. Voprosy Jazykoznanija 5 : 5. 55-67 /1956/ .

Mathesius, V. O tak zvaném aktuálním členení vetnem: Čestina a obecný jazykozpyt. Praha, 1947. 234-42.

-- Zakladní funkce českého pořadku slov. Čestina ... 327-52.

Melničuk, A.S. Porjadok slov i sintagmatičeskoje členenije predloženij v slavjanskich jazykach. K. 1958.

Simonyi, Zs., A magyar szórend. Magyar Nyelvőr XXXI. 57-61, 121-9, 18o-6, 234-41, 289-3o1, 359-79, 424-41, 437-96 /19o2/

Skalička, Vl. La fonction de l'ordre des éléments linguistiques. TCLP 6. 129-33 /1936/ .

Travniček, F., Základy československého slovosledu. Slovo a slovesnost 3. 78 /1937/ .

Worth, D.S., Ob otobraženii linejnych otnošenij v poroždajuščich modeljach jazyka. Voprosy Jazykoznanija 13: 5. 46-58 /1964 / .

# SOME QUESTIONS OF PHRASE STRUCTURE GRAMMARS I.

by S. Abraham
Computing Centre of the Hungarian
Academy of Sciences

0. In the last years increasing attention has been paid to generative grammars. Not all the generative grammars were submitted to a study of equal intensity, but special interest was shown towards phrase structure grammars /PSG/.

1. A generative grammar /G-grammar/ is a quadruple

$$/1/ \qquad G = /V, V_t, \Sigma, F/$$

where V is a finite set of symbols /letters/called dictionary, $V_t$ is a /proper/ subset of V called terminal dictionary. The elements of $V_t$ are called terminal letters. Over V the /associative and non-commutative/ operation of concatenation is defined /noted with /. In the following the sign of this operation will not be noted /so instead of X Y we shall write XY/. The operation of concatenation yields the set of sequences /or words/ over V. The null-sequence, noted by E is also considered /for E we have E X = X E = X for every X V/. $\Sigma$ is a finite set of sequences over V called initial sequences. F is a finite set of pairs $(\varphi, \psi)$, where $\varphi$ and $\psi$ are sequences over V, called rules /of the grammar/. $\varphi$ is called the left-side of the rule and $\psi$ its right-side. To apply a rule f:$(\varphi, \psi)$to a sequence x means to replace in x the first occurance of $\varphi$ with $\psi$. If x does not contain any occurance of the left-side of a given rule f, we say that f cannot be applied to this x. If by applying f to x we obtain x' we shall write /f[x]= x' or x x' and say that x directly generates x'.

The sequence of words

$$/2/ \qquad x_1, \dots, x_n$$

is called a derivation in G if every $x_i$ $/1< i \leq n/$ is obtained by applying a rule from F to $x_{i-1}$. In this case we say that every $x_i$ $/1< i \leq n/$ can be generated by any $x_j$, if $i<j$, and denote this by $x_j \Rightarrow x_i$. The derivation /2/ is a terminated derivation if to $x_n$ no rule from F can

be applied. In this case we say that $x_1$ is the first word of the derivation and $x_n$ its terminated word. If /2/ is a terminated derivation and $x_n$ is a sequence over $V_t$, we call it a terminal derivation and $x_n$ its terminal word. The set of all the terminal words of the /terminal/ derivation with the first words from $\Sigma$ /in this case called base words of the derivation/, forms the language L generated by (G) /denoted/ as L G /.

A grammar G is called a phrase structure grammar /PSG/ if the rules $f_i$ from F are /all/ of the form

/3/ $(\varphi_i A \psi_i \ , \ \varphi_i \omega \psi_i)$

where (a) $\varphi_i, \omega, \psi_i$ are sequences over V, (b) $\omega \neq E$ (c) A V.

If for $f_i$ we have

/4/ $\varphi_i = \psi_i = E$

$f_i$ is called a context-free rule /otherwise $f_i$ is called a context-restricted rule/. The PSG is called a context-free grammar if all its rules are context-free rules. Otherwise we have a context restricted grammar /briefly, CF and CR grammars respectively/.

That the CR grammars are essentially different from the CF grammars follows from the /known/ fact that all CF grammars are CR grammars, but not conversely. A well-known example of a CR language /i.e. language generated by a CR grammar/ but not CF language is

/5/ $L = \{a^n \ b^n \ c^n\}$

At a rather early stage of the study of PS grammars the /important/ problem of the closedness of these grammars under different operations was considered. The problem is that if $L_i$ is a language generated by a grammar $G_1$ and $L_2$ by a grammar $G_2$, then the language $L_1{}^{*} L_2$ /where $*$ denotes the considered operation/ is of the type of $G_1$ or of $G_2$ or of neither of them. The operation considered was "product", "reunion", "intersection" and "complementation". The definition of those operations and the main results can be found in [1], [3].

Another very important question is that of establishing whether some important properties are decidable or not for these classes of grammars. A table containing the result claimed to have been obtained up till the present can be found in [6]. I will here only set out those properties in which these two grammars are said to differ:

| Property | CFG | CRG |
|---|---|---|
| 1. is the language generated by a grammar infinite? | D | U |
| 2. is the language generated by a grammar infinite? | D | U |
| 3. For any strings $\phi, \psi$ can some string including $\psi$ be derived from $\phi$ in a grammar? | D | U |

where D indicates that the property in question is decidable, U that it is undecidable.

These results /and especially some considerations derived from the undecidability of the property 3/ make doubtful the fitness of context restricted grammars for modeling natural languages. On the other hand, the context free grammars turned out to be generatively inadequate for this task. One possible way out is to construct grammars more powerful than the context free ones but in which the mentioned properties are decidable. I shall follow this way by constructing such a grammar, but it will turn out that even for context restricted grammars the above mentioned three properties are decidable.

2. In the following I shall introduce a new type of generative grammar, called <u>matrix grammar</u>.

A <u>matrix grammar</u> is a quintuple

$$/6/ \qquad G = (V, V_t, \Sigma, F, F^*)$$

where

$$\bar{G} = (V, V_t \ \Sigma, F)$$

is a context free grammar as defined above. $F^*$ is a finite set of matrices /called <u>matrix rules</u>/, defined as follows:

/i/ $f^*$ is a matrix rule if it has the form

$$\begin{bmatrix} f_1 \\ \cdot \\ \cdot \\ \cdot \\ f_n \end{bmatrix}$$

with $f_i$ F /not necessarily $f_i \neq f_j$/.

/ii/ $f^*$ is a matrix rule if it has the form

$$\begin{bmatrix} f_1^* \\ \circ \\ \cdot \\ \cdot \\ f_n^* \end{bmatrix}$$

where $f_i^*$ are matrix rules or belong to F.

To apply a matrix rule a $f^*$ to a string x means to apply all the context free rules which form it, to x, in the given order. If by applying $f^*$ to x we obtain x', then we write $f^*[x]= x'$ and /or $x \xrightarrow{*} x$/ say that x directly generates x'.

The sequence of words

/7/ $x_1, \dots, x_n$

is called a derivation in G if every $x_i (1< i \leq n)$ is obtained by applying a rule from $F^*$ to $x_{i1}$. In this case we say that every $x_i$ is generated by any $x_j$ if $i < j$ and denote this by $x_i^* \overset{*}{\Longrightarrow} x$.

/7/ is a terminated derivation if no rule from $F^*$ can be applied to $x_n$ / and we say that $x_1$ is the first word of the derivation and $x_n$ is its terminated word/.

If /7/ is a terminated derivation and $x_n$ is a sequence over $V_t$, we call it a terminal derivation and $x_n$ its terminal word. The set of all the terminal words of the terminal derivations with the first word from $\Sigma$, forms the language generated by G /L (G)/.

Before turning to the more detailed study of matrix grammars let us mention that if a matrix rule contains matrix rules, the sign of the inner matrices can be dropped. A matrix rule in which the signs of the inner matrices are dropped is called a normal matrix rule and the

grammar which contains only normal matrix rules a <u>normal</u> matrix grammar. As to each matrix grammar a strongly equivalent normal matrix grammar corresponds, by the term matrix grammar I shall usually mean a normal matrix grammar /and the deletion of the inner matrix signs will be called <u>normalization</u> of the given matrix rule/.
The main theorem of the present paper is the following:

<u>Theorem.</u> For every given context-sensitive grammar G a strongly equivalent matrix grammar $G_M$ can be constructed.

Without any loss of generality we can consider that a CR grammar contains also context-free rules and that in the list F of the rules of the grammar we have first listed the context-free rules followed by the context restricted ones.

Now let us proceed to the proof of our theorem.

First let us suppose that the context-restricted grammar G for which we shall construct the equivalent matrix grammar, contains only one context restricted rule, say

$$f : (x = \varphi A \psi , \varphi \omega \psi)$$

Now, consider the grammar $\bar{G}$, which contains only the context-free rules of G. If in $\bar{G}$ a string

$$w = \ldots x \ldots$$

is derivable /this problem is decidable/ then we have a string w' and a rule $f_j$ so that

$$f_j (w') = w , \qquad \text{so } w' \to w$$

Without any loss of generality / as the context-free rules are not ordered / we can suppose that $x \notin w'$. Then we have a string $x' \subset w'$ so that

$$x' \to uxv$$

where u,v are /possibly null/ strings over V, and

$$\ell (x') \leq \ell (x) = m$$

where $\ell(x)$ the <u>length</u> of x, is the number of the symbols which form x.

As the number of all the strings y with the property $\ell (y) \leq m$ is finite, and their set can be effectively constructed, we can also effectively construct the set

$$M = \{ x' \mid \ell (x) \leq m \ \& \ (\exists u)(\exists v) x' \to u x v \}$$

Now let us have the derivation

$$x_o \longrightarrow x_1 \longrightarrow x_2 \longrightarrow \ldots \longrightarrow x_n = x' \longrightarrow x$$

where $x_o \in \Sigma$ and $f_i(x_{i-1}) = x_i$ $/1 \leq i \leq n/$.

In the sequence of rules

$$f_1, \ldots\ldots, f_n$$

one and only one, say $f_k$, will introduce the A of $\varphi$ A $\psi$. Then we form the matrix rule

$$f^{*} : \begin{bmatrix} \bar{f}_k \\ f_{k+1} \\ \cdot \\ \cdot \\ \cdot \\ f_j \end{bmatrix}$$

where $\bar{f}_k$ differs from $f_k$ only by replacing in it A with $\bar{A}$ $/\bar{A} \notin V/$. Now if we add to the rules of $\bar{G}$ the matrix rule $f^{*}$ and the context-free rules $/\bar{A}, \bar{\omega}/$ and $/\bar{A}, A/$ /where $\bar{\omega}$ is $\omega$ with A, if any, replaced by $\bar{A}$/ we obtain a grammar $G^{*}$ so that

$$G = G^{*}$$

From this, by induction /on the number of the context-restricted rules a context-restricted grammar contains/ and the remark that no matrix rule of the type

$$f^{*}: \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ f^{*} \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}$$

will arise during the constructions, follows the theorem.

Evidently, any context-free grammar G with the rules F: $\{f_1, \ldots, f_n\}$ can be regarded as a matrix grammar G, with $F^{*} : \{[f_1], \ldots, [f_n]\}$.

For the non-context-free language

$$L = \{a^n b^n c^n\}$$

we have the following matrix grammar that generates it:

$$G = (V, V_t, \Sigma, F, F^{*})$$

with

$V = \{S, X, Y, Z, a, b, c,\}$ $\qquad V_t = \{a, b, c\}$

$F: \{(S, abc)\ ,\ (S, aXbYcZ)\ ,\ (X, aX)\ ,\ (Y, bY)\ ,$
$\quad (Z, cZ)\ ,\ (X, y)\ ,\ (Y, b)\ ,\ (Z, c)\}$

$F^{*}:$

$$[S,\ abc]$$

$$[S,\ aXbYcZ]$$

$$\begin{bmatrix} X, & aX \\ Y, & bY \\ Z, & cZ \end{bmatrix}$$

$$\begin{bmatrix} X, & a \\ Y, & b \\ Z, & c \end{bmatrix}$$

3. As the proofs given in [1] for the decidability of the properties 1, 2, 3 /Table 1/ for context-free grammars remain true if instead of context-free rules we consider /normal/ matrix rules, these properties are also decidable for the matrix grammars, and this is not suprising, as in a concrete derivation, using only context-free rules, it is irrelevant whether these rules /or parts of them/ form a matrix rule or not.

Let us give proof only of the most important of them:
Given a matrix grammar G and the strings $\varphi, \psi$

$$/l\ (\varphi) \leq l\ (\psi)\ = m/$$

it is decidable whether

$\phi \xrightarrow{*} u\psi v$ /where u,v are possible null strings over V/

Let us construct /following [1] / inductively the sets

$$A_m^1 = \{z \mid \ell(z) \leq m \ \&\ (\exists u)(\exists v)(A \xrightarrow{*} uzv)\ \&\ A \in V\}$$

$$A_m^k = \{z \mid \ell(z) \leq m \ \&\ (z \in A_m^{k-1} \vee (\exists u)(\exists v)(\exists y)(y \in A_m^{k-1}\ \&\ y \xrightarrow{*} uzv)\}$$

All of them can be effectively constructed.

Evidently $A_m^{k-1}$ $\quad A_m^k$ and if $A_m^{k-1} = A_m^k$ ,
then $A_m^k = A_m^{k+1}$. So the sequence $A_m^1$ , $A_m^2$ , . . . strictly increase until an equality is obtained. As the total number of strings not longer than m is finite, we have a $\beta$ so that

$$\lim_{k \to \infty} A_m^k = A_m^\beta \quad .$$

As, if $\phi \overset{*}{\Rightarrow} u \Psi v$ then $\phi \in A_m^\beta$, this can be effectively tested.

The erroneousness of the "proof" of the undecidability of property three for context-restricted grammars given in [3] can be easily shown. As to any context restricted grammar an equivalent matrix grammar corresponds the decidability of properties 1,2,3 follows from their decidability for the matrix grammars.

4. It is often argued that for the description of natural languages the PS grammars are not satisfactory. This problem can be regarded from two points of view:

/i/ given a natural language L, there can be constructed a PS grammar so that $L(G) = L$ ? This is the so-called generative adequacy problem.

/ii/ if /i/ is answered positively, then there exists a PS grammar which generates L and assigns to all the generated sentences a correct structure? This is the so-called explicative adequacy problem.

If the generative adequacy problem can be /and is/ formulated precisely /and formally/ in a satisfactory way, with regard to the problem of explicative adequacy this can not be done so easily, as the notion of a "correct structure" is not defined satisfactorily. To this question I shall return later.

With regard to the generative adequacy of PS grammars, no proof is given in any sense; neither that they are generatively adequate, nor the contrary is proven. But the problem of the explicative adequacy is answered in the negative /we shall see that without any serious grounds/ and new rules are therefore introduced, the so-called /singular/transformational rules, and the transformational grammars are considered, which consist of two components, a PS component /in fact a PS grammar/ and a transformational component, which acts in the output of the PS component. The more or less exact definition of transformational grammars can be found in [5] .

In the following I shall argue that the PS grammars are in exactly the same way generatively adequate as the transformational grammars are.* In other words, the PS component can be so extended that it generated all the sentences which are generated by the transformational grammar. Finally I shall discuss in detail the problem of the explicative adequacy of the PS grammars. These problems will be the subject of the second part of the present paper.

---

* This result was first stated in the paper submitted to the Conference on Mathematical Linguistics held in Bucarest /21-25 October, 1963/. Unfortunately, the report on the proceedings of this conference, given by A. Rocerik-Alexandrescu /in Studii si Cercetari de Lingvistica, XV, 1964 /1/, p. lol, Bucarest/ mentions only the /transparent/ starting definition of generative models of languages, omitting thus the essential statements of the paper.

BIBLIOGRAPHY

[1] Y.,Bar-Hillel, M. Perles, E. Shamir, "On formal properties of simple phrase structure grammars", Applied Logic Branch, Hebrew Univ. of Jerusalem, Tech. Rept. No 4; 196o.

[2] N.,Chomsky, "On certain formal properties of grammars", Inf. and Control, vol. 2 pp. 137-167; June, 1959.

[3] N.,Chomsky, "Formal Properties of Grammars" in "Handbook of Mathematical Psychology", John Wiley and Sons, Inc., New York, N.Y. vol. 2, pp. 323-418; 1963.

[4] N., Chomsky, M.P., Schützenberger, "The Algebraic Theory of Context Free Languages in Computer Programming and Formal Systems", North-Holland, Amsterdam, pp. 118-161; 1963.

[5] N.,Chomsky, On the Notion "Rule of Grammar" in Proc. of Symposia in Appl. Math., vol. xii Amer.Math.Soc., Prov. R.I., pp.6-25; 1961.

[6] P.S., Landweber, "Decision Problems of Phrase-Structure Grammars", IEE Trans., vol. EC-13, pp. 354-462; 1964.

[7] G.H., Mathews, "Discontinuity and asymmetry in phrase structure grammars", Inf. and Control, vol. 6, pp. 137-146, June, 1963.

# SOME QUESTIONS OF SEMANTIC THEORY

F. Kiefer

1. Semantic theory plays only an interpretative role in linguistic theory, i.e. it takes the sentences for granted and sets out to do no more than answer certain questions raised with respect to a given sentence or sentences. However, it should be made clear that it makes no sense to distinguish between grammar and semantics if we do not assume that there exist two disjunct sets of categories, namely that of grammatical and that of semantic categories. The notion of category is taken as a primitive notion in the following, but our above statement would be made clearer by saying that if a morpheme $w_i$ is specified by a category $k_i$, then there must be at least one other morpheme $w_j$ that cannot be specified by this category. To put it differently, categories play a distinctive role, they enable the differentiation of morphemes from one another. Grammar is undoubtedly based on a set of such categories, and semantics cannot be considered as a separate theory if we cannot tell semantic categories from grammatical ones. Therefore we will assume the existence of the two disjunct sets of categories. [1]

Now semantic theory can be conceived as a pair

$$\langle V , R \rangle$$

where V stands for the vocabulary and R for the set of semantic rules. [2]

2. Each vocabulary item consists of two parts: of a lexical morpheme and of a set of categories. The set of categories contains grammatical and semantic categories. The grammatical categories are part-of-speech categories, the so-called functional categories can be introduced only after the actual generation of the morpheme. [3] The semantic categories provide the semantic characterization of the given morpheme.

Let us denote the set of all grammatical categories required by the grammatical characterization of the morphemes by $K_G$ and the set of all semantic categories required by the semantic characterization of the morphemes by $K_S$. Both sets can be divided into classes such that

$$k_{u_i} < k_{u_{i+1}} < \dots < k_{u_{i+r}} \qquad /1/$$

for the categories belonging to any class, where r stands for the number of categories in a given class and the relation $<$ means that if $k_i$ is a category of the morpheme w and

$$k_i < k_j$$

then the morpheme w may be characterized by the category $k_j$ as well. Of course, it may happen that some classes contain only a single element.

Let us call the relation $<$ linear subordination. Let us denote the classes obtained by linear subordination by $C_1, C_2, \dots, C_p$.

The set of categories assigned to a given morpheme /grammatical as well as semantic categories/ is called the distinctive set of the given morpheme. The distinctive set can be made redundancy-free in the following way. If the distinctive set contains more than one category of a class $C_i$, then only the leftmost element of these categories is taken where the meaning of "leftmost" will be obvious if we observe that the categories of the distinctive set form a substring of /1/.

For reasons of economy it is advisable to consider always a redundancy-free distinctive set as the second part of the vocabulary items.[4]

Each category if necessary may be positively or negatively specified with respect to a given morpheme.

3. The general form of the semantic rules is

$$\varphi \rightarrow \psi$$

where $\longrightarrow$ means: rewrite and $\varphi, \psi$ will be specified below.

Let us denote two morphemes by $w_1$ and $w_2$. The distinctive set of $w_1$ is denoted by $K_1$ and that of $w_2$ by $K_2$.[5] Further, let us denote a distinguished set of categories by $M_1$ and another one by $M_2$.[6] Lastly, let us denote the semantic rules by $f_1, f_2, \dots, f_s$. The necessary and sufficient condition for the applicability of a rule $f_i$ can be given by the following conditions

$$M_1 \subseteq K_1 \qquad \text{and} \qquad M_2 \subseteq K_2 \qquad /2/$$

where it is assumed that the semantic rules apply always to two morphemes at once.

To apply an $f_i$ rule means that first the matrices corresponding to $K_1$ and $K_2$ are constructed and denoted by $m_1$ and $m_2$ respectively. The matrix representation of category sequences is quite trivial therefore I do not discuss this topic in more detail here.

Now the semantic rules are of the following form

$$m_1, m_2 \longrightarrow m$$

i.e. two matrices are replaced by one matrix if the conditions /2/ are met.

Thus, the semantic rules are given by matrix mapping. We can impose some formal restrictions on this mapping. Let be the type of the matrix $m_1$ $p \times q$ and that of $m_2$ $p' \times q'$. Then, the numbers of rows in m will be precisely p+p', the numbers of columns, however, at least v = max /q,q'/. The number of columns will be precisely v if and only if

$$K_1 \subseteq K_2 \qquad /3/$$

As in general the condition /3/ is not fulfilled we may obtain even $K_1 \cap K_2 = 0$. In this case the matrix m has precisely q + q' columns. As in the general case the number of columns is between v and q + q', the matrix m must be normalized, that is it has to be extended in a way that the corresponding categories belong to the same column. [7] Then, the free places are filled in with 0's.

It is easy to give a general procedure for the normalization of matrices in the above sense however complicated the matrices are.

Now it is clear that any semantic rule $f_i$ is given by the conditions $/M_i, M_j/$.

4. Let us denote two strings by a, b. These strings are said to be <u>compatible</u> if and only if there exists at least one semantic rule $f_i$ such that $f_i$ contains on its left-hand side the matrices corresponding to a and b in this order /i.e. not in the order corresponding to /b,a/. The compatibility relation is denoted simply by parantheses. If a,b, are compatible then we write /a,b/.

Three strings, a,b, and c, are said to be compatible, if and only if, either

$$//a,b/,c/ \text{ or } /a,/b,c// .$$

In general the strings $a_1,a_2,\ldots,a_n$ are said to be compatible, if and only if, either

$$//a_1,a_2,\ldots,a_{n-1}/,a_n/$$

or

$$/a_1,/a_2,a_3,\ldots,a_n//,$$

where $n \geq 2$.

If /a,b/ then in general /b,a/ does not hold, and if /a,b/ and /b,c/ then /a,c/ need not be necessarily true. In other words, the compatibility relation is neither commutative nor transitive and is consequently not an equivalence relation.

Let us denote by a given sentence generated previously by the grammar G. The grammar G provides for the sentence s a structural description that is generally referred to as P-marker. Each P-marker is equivalent to a labeled bracketing. There are many linguistic facts that suggest that the semantic interpretation should be based on P-markers rather than on terminal strings /sentences/. It is clear, for instance, that it must be decided whether a given constituent is compatible and on the other hand, there is no need of examining the compatibility of strings that do not form a constituent. The constituents determine the starting point and the direction of the compatibility proof. These and other facts suggest the following definition:[8]

The sentence

$$a_{x_1}a_{x_2}\ldots a_{x_r}$$

is called <u>meaningful</u> if and only if all its constituent strings given by the labeled bracketing are compatible.

In the case of meaningful sentences it is possible to obtain a single matrix by applying semantic rules. This matrix is called <u>distinctive semantic matrix</u>.

It is clear, then, that a sentence s has so many meanings as distinctive semantic matrices. If s has i different distinctive semantic matrices then we say that it is <u>i-ways ambiguous</u>.

If it has only one distinctive semantic matrix then we say that the sentence s is <u>semantically unambiguous</u>.

If the sentence has no distinctive semantic matrix then we say that it is <u>semantically anomalous</u>.

It is apparent that two sentences will have as many meanings in common as they do dinstinctive semantic matrices.

If two sentences have i distinctive semantic matrices in common then we say that an <u>i-ways synonymy</u> holds between them.

If two sentences have each exactly i dinstinctive semantic matrices and if between them an i-ways synonymy holds, the two sentences are <u>fully synonymous</u>.

The full synonymy is, incidentally, an equivalence relation.

5. As pointed out above compatibility is not an equivalence relation. However, it is easy to define on the basis of compatibility a relation which is an equivalence relation.

Two words, a and b, are called equivalent if /a,y/ and /b,y/ hold true at the same time.

By denoting this equivalence by $\Rightarrow$, we write

$$a \Rightarrow b \ .$$

The defined relation is an equivalence relation because /i/ $a \Rightarrow a$ holds. Namely, if /a,y/, then evidently also /a,y/.

/ii/ if $a \Rightarrow b$, then $b \Rightarrow a$. Namely if /a,y/ and /b,y/ then also /b,y/ and a,y/.

/iii/ if $a \Rightarrow b$ and $b \Rightarrow c$, then $a \Rightarrow c$. Namely, if /a,y/, /b,y/ and /c,y/, then also /a,y/, /c,y/.

On the basis of this equivalence it is possible to define a partition of the set of morphemes compatible with y:

$$/X,y/$$

where X contains all the morphemes being compatible with y. In this case y is called the <u>generator element</u> and X the <u>semantic range</u> of y. If two different generator elements, y, and z, have the same semantic range, then they are <u>semantically similar</u>.

Let us divide the set of morphemes of a given language L according to their part-of-speech categories. This way we obtain the classes

$D_1, D_2, \ldots, D_n$ which are called in accordance with the usual terminology the distributional classes of L. Each $D_i$ contains a finite number of morphemes.

Let two classes X and X' be defined by

$$/X,y/ \quad \text{and} \quad /X',y'/.$$

Now we consider the set $\bar{X}$ defined by

$$\bar{X} = X \cap X'.$$

We claim that the morpheme y is characterized with respect to the morpheme y' by the set $X^+$ for which

$$X^+ = X - \bar{X}.$$

And, on the other hand, the morpheme y' is characterized with respect to the morpheme y by the set $X^{++}$ for which we have

$$X^{++} = X' - \bar{X}.$$

However, there are further possibilities of formalizing the linguistic concept of semantic relationship between morphemes. Let $D_k$ be any distributional class of L. Let us divide $D_k$ according to some semantic categories into the subclasses [9]

$$\Sigma_1, \Sigma_2, \ldots, \Sigma_r$$

We want to answer the following question: what can be said about a class $\Sigma_k$ from a semantic point of view?

We claim that the subclass $\Sigma_k$ is semantically characterized by the set X'' defined in the following way

$$X'' = X_1 \cap X_2 \cap \ldots \cap X_s$$

where

$$/X_1,y_1/, /X_2,y_2/, \ldots, /X_s,y_s/$$

and $y_i \in \Sigma_k$ for every $i=1,2,\ldots,s$. The partition must be defined in a way that

$$X'' = X_1 \cup X_2 \cup \ldots \cup X_s \,.$$

Or, to put it differently, that

$$X_i - X'' = 0$$

for every $i=1,2,\ldots,s$.

In this case the conditions /2/ may be established in the following way. $M_i$ contains the semantic categories that characterize the class $X''$ and $M_j$ the semantic categories of $\Sigma_k$.

If the condition

$$X_i - X'' = 0$$

is not met, then this is due to some exceptional elements which, however, can be ruled out by handling them separately.

6. To sum up, it is hoped that I have succeeded in pointing out the possibility of formalizing some linguistic notions about semantics, a fact, that may be of interest both in theory and in the practice of exact linguistic theory. I do not claim, however, that all the problems touched upon in the course of the discussion have been solved satisfactorily. The present paper should be considered no more than a tentative contribution to a formal semantic theory.[10]

Notes and references

1 A more detailed discussion of this topic is to be found in Kiefer,F., Abraham,S., Some Problems of Formalization in Linguistics, Linguistics /forthcoming/.

2 The structure of semantic theory has first been tackled by Katz and Fodor /Katz,J.J., Fodor,J.A., The Structure of a Semantic Theory, Language 39 /1963/, pp. 17o-21o and by Katz and Postal /Katz, J.J., Postal, P.M., An Integrated Theory of Linguistic Descriptions, The M.I.T. Press, Cambridge /Mass./, 1964/. Another approach has been worked out by Abraham and Kiefer /Abraham, S., Kiefer,F., A Theory of Structural Semantics, Mouton and Co., The Hague /forthcoming/, Kiefer,F., About the Formalization of Semantic Relations /in Hungarian/, Általános Nyelvészeti Tanulmányok /Studies in General Linguistics, ed. Zs.Telegdi, forthcoming//. Some of the ideas first presented in the above monographs appear here in a slightly revised form with some important extensions.

3 The question how these categories can be introduced is tackled in some detail in the monographs mentioned in 2 .

4 General rules for the expansion of the redundancyfree distinctive set can be established in a rather trivial way, so I do not want enter into this question here.

5 Notice that $K_1$ and $K_2$ are not redundancy-free: This is because the application of rules generally requires the expanded set of categories.

6 The sets $M_1$ and $M_2$ will be specified below.

7 This could be made more precise only by defining first the correspondence between the set of matrices and the subsets of categories.

8 Cf. Katz and Fodor , Katz and Postal in 2 and Chomsky,N., Categories and Relations in Syntactic Theory /mimeographed/, Cambridge /Mass./, M.I.T. 1964.

9 This partition could only be defined precisely in the case of a concrete language, here the main points of such a definition are given.

10 The formal definitional apparatus given in this paper is exemplified in the two monographs quoted in 2 .

# THE ANALYSIS OF PREPOSITIONAL CONSTRUCTIONS

Gy. Sipőczy

## I. The theoretical basis of our method

At the conference on "Mathematical linguistics and machine translation" held at the Hungarian Academy of Sciences on March 8-9, 1962 we put forth a method the application of which makes possible - in the course of the mechanical analysis - the analysis of the attributive or adverbial role of the prepositional constructions. [1.] The main points of our proposal were as follows.

In the sentence the prepositional construction following the substantive /or in the case of genitive relation the nominal group/ may be the attribute of the preceding substantive /in the case of a nominal group one substantive of the group/ or the adverb of the verb. If the prepositional construction follows an adjective then it is either the adverb of the preceding adjective or the adverb of the verb. In all those cases when there is a substantive or an adjective before the prepositional construction, we have to find out with which word of the sentence the prepositional construction forms a syntagma.

To some verbs, substantives and adjectives the substantives may be connected only by some prepositions as adverbial or adjective complements. These prepositions are characteristic of the verbs, substantives and adjectives /i.e. those prepositions with the help of which the latter may form a syntagma/, moreover, it is also characteristic with what frequency a preposition occurs as a required postposition of a given morpheme. We may group the verbal, substantival and adjectival morphemes in a way that those of them will belong to the same group which form a syntagma by the aid of the same preposition. By classifying the morphemes we have to take into consideration the rank of frequency of the occuring prepositions: within each

---

1. The system is based on the algorithm elaborated by Gy. Hell where the first code number /K.I./ indicates the part-of-speech properties and the second one /K.II./ the morphological properties.

group we rank the prepositions according to their decreasing frequency. The serial numbers of the groups obtained in this way are stored as code numbers /K.III./ adjoining the morphemes under consideration. We use of these code numbers in the subprograms of the prepositions for the determination of the syntactic role of the prepositional constructions. The code numbers are placed in these subprograms in a way that it should be clear in which group the given preposition occurs first, second, third etc. So, for instance, if the prepositions stands in the sentence after the substantive, the computer has only to find out on the basis of the subprogram whether the given preposition occurs first in the group of the preceding substantive or in the group of the verb. If in the former then the program considers the prepositional construction as attribute, if in the latter as adverb.

The method outlined above facilitates decisions about the syntactic role of the prepositional construction on the basis of statistical considerations. This fact is undoubtedly disadvantageous because the results obtained in this way are not always reliable, but so long as we have no suitable semantic system at our disposal which could be taken as a firm basis for the decision - we have to content ourselves with this provisional solution. For our purposes the question may be raised in the following way:

1./ Can the grouping of the verbal, substantival and attributive morphemes outlined above yield a firm basis for the determination of the syntactic role of the prepositional constructions?

2./ Does the quantity /length/ of the text on the basis of which the classification /grouping/ was carried out exert any influence on the results?

II. <u>The practical control of the method</u>

To decide about the practical applicability of the method we took a sample from the book:

Ю.В. Новиков: Элементарные частицы. Москва, 1963.

The method was checked against the prepositional constructions found in the Introduction and the First Chapter /49 pages altogether/ of the above-mentioned book.

The clarification of the syntactic role of the syntactic constructions called for the grouping of the verbs, substantives and adjectives according to the required postpositions. First I considered only the morphemes that require postpositions occuring in the First Chapter and the Introduction, thereafter - to find out how the length of the text alters the result of the analysis - the morphemes that require postpositions of the Second Chapter were added and then the prepositional constructions of the Introduction and the First Chapter were analysed by the aid of the groups obtained on the basis of the Introduction, the First and Scond Chapters /7o pages altogether/.

1. The number of morphemes:

The number of the morphemes that require postpositions and that are of account with respect to the clarification of the syntactic role of the prepositional constructions is given by the following table:

| | Introduction + Chap.I. | Intr. + Chaps.I,II. |
|---|---|---|
| Verb: | 197 | 274 |
| Substantive: | 1o7 | 161 |
| Adjective: | 38 | 58 |

2. The number of groups

According to the classification described in I. the set of all morphemes can be divided into 52 groups. It is clear, however, that the more groups we have the more accurate our results are. As in the code-system that underlies our examinations the part-of-speech properties of the different morphemes are labeled by special code numbers /K.I./, I divided the morphemes - departing from the original idea - according to their part-of-speech category, i.e. I grouped the verbs, substantives and adjectives separately. This way I obtained 3 times 64 groups in the system of 6 bits, i.e. the treble of the original number. Presumably the examination of a much larger corpus /several hundreds or thousands pages/ would require an even greater number of groups. In this case even the code system would have to be considerably altered. On the basis of the examined text we obtained the following number of groups:

| | Intr. + Chap.I. | Intr. + Chaps.I,II |
|---|---|---|
| Verb: | 43 | 45 |
| Substantive: | 36 | 36 |
| Adjective: | 16 | 19 |

This comparison reveals that there is no considerable increase of groups when the length of the analyzed text is enlarged.

3. The principles of the grouping

a./ In the case of prepositions requiring more than one case it is necessary to differenciate the different governments /B + Acc; B + Pr. case/.[+] The morphological analysis makes for this distinction. The prepositions with more than one government are analyzed by means of different subalgorithms each corresponding to a given government. In addition to this semantic restrictions are needed as well /preposition + substantive marking measure, preposition + time substantive, etc./.

b./ In the different groups the place of each preposition is determined by the frequency with which it occurred as a required postposition of the morphemes belonging to the given group. In the case of verbs and substantives the first place is occupied by those prepositions that occur as strong requirement and those that occurred lo times or more. The second place is taken up by the prepositions occuring 9 times and so forth. "Strong requirement is defined by Yordanskaya in the following way: Strong requirement means the property of stems on the basis of which it is possible to guess - with great enough likelihood - one or more word forms depending on them". In the groups of verbs and substantives the prepositions occurred in the following places:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Serial number: | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | lo. |
| Number of occurrences: | Strong requirement or lo occurrences | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |

In the case of adjectives this grouping cannot be maintained because of the low frequency of their occurrence which causes a considerable distortion in the result. In the corpus under consideration it often happens that the adjective occurs in a syntagma only once or twice with a preposition that - in another text or sample - is its almost constant or at least very frequent requirement while the same preposition occurs in the second or third place in some verbal constructions. This would lead to the result that the prepositional construction is - contrary to fact - the adverb of the verb. As a consequence of the above considerations I have connected two serial numbers with one occurrence in the case of adjectives. This modific-

---

[+] Pr = prepositional case

ation compensates to a considerable extent the distortion caused by the low frequency of occurrence. In the groups of adjectives we therefore obtain the following table:

| | | | | | | |
|---|---|---|---|---|---|---|
| Serial number: | 1. | 2. | 3.4. | 5.6. | 7.8. | 9.1o. |
| Number of occurrences | Strong requirement | 5 or more | 4 | 3 | 2 | 1 |

The disproportionateness of the number of occurrences at the cost of the adjectives holds in the case of any text no matter how large it is. To avoid that this disproportionateness do not bear on the result unfavourably we have to take into consideration the ratio of the morphemes in the course of grouping. This may be carried out in the simplest way as follows: We compute how many times the number of verbs and substantives is bigger than the number of adjectives in the given corpus. Now we multiply by this number the number of occurrences of the preposition required by the adjective and rank it into the class corresponding to the obtained product. A similar procedure can be used to eliminate the disproportionateness between verbs and substantives. Let us assume, for instance, that the ratio of verbs, substantives and adjectives is 5 : 4 : 1. If, for instance, a preposition occurs 15 times after a verb and only 3 times after an adjective in the corpus under consideration then the two cases belong to the same class.

As the place of the prepositions belonging to the different groups depends on their number of frequency it is natural that we have also empty places and that in the case of identical numbers of frequency we can have more than one preposition in the same group /under the same serial number/. So, for instance, we have the following prepositions belonging to the first group of verbs /the serial number of the preposition stands in parantheses/:

В +Acc. /1/;В+Pr. /4/; при /6/; на +Acc. /9/; В+Acc. /idő/ /1o/; среди + Gen. /1o/.

c./ Since the examination aims only at the clarification of the syntactic role of the prepositional constructions the case government was left out of consideration.

d./ The active and passive verb morphemes - as they require the same prepositions in general - are considered as belonging to the same group. However, if the required preposition is different in the case of a passive morpheme from the active one /in this case even the meaning will be different / it is put into another group. So, for instance, the verb "заключать" may require the following prepositions: из ; В +Acc; о+Pr; для ; the verb "заключаться" : В+Pr; В+Acc. We have a similar case with verbs such as : иметь, иметься , etc.

e./ Where the preposition occurs as a strong requirement it is put in the first place even in the case when the number of occurrence is small or another preposition is required more frequently. So, for instance, the verb "встречаться" occurs with the preposition 6 times and with "В+Pr" 9 times and yet "с" will be put first in the list.

f./ Those morphemes that have more than one strong requirement correspondingly occur in several groups. This results in a semantic classification at the same time because the meaning of the morpheme alters in the case of different strong requirements. E.g. the verb "состоять" with the strong requirement "из" /to consist of/ belongs to the fourth and with the requirement "В+Pr." /to consist in/ to the second verbal group.

g./ In the case of compared adjectives the required postposition is generally different from that required by the positive degree, consequently they must be handled as separate morphemes. We encountered, for instance, the prepositions "В+Pr." and на +Pr." after "малый" in the given corpus and "В+Acc" /measure/ and на +Acc" measure/ after "меньше" but "из"after "наименьший".

h./ Those prepositional constructions that play the role of idioms, i.e. occur always in the same form and have prepositional, adverbial or noun-suffixal function, must be handled separately as well. Such constructions are e.g. в зависимости от, в качестве, в конце концов, в отличие от, в самом деле, в связи с, в силу, вместе с, за счет, наряду с, по отношению к, по сравнению с, с помощью, с точки зрения, etc.

3. The application of the grouping method

The groups obtained according to the above procedure enable the determination of the syntactic role of the prepositional constructions in the following way:

a./ First, we have to find out to which group the substantive /substantives/ or adjective preceding the prepositional construction and the verb of the sentence belong /the number of the group equals the code numbers of K.III./.

b./ We state the serial number of the prepositions required by the morphemes in the groups to which the corresponding morphemes belong. The prepositional construction will be required by the morphemes in the group whose serial number the lowest is /lower serial number indicates more frequent occurrence/.

Some simple examples:

a./ ..., нужной для разделения молекулы на отдельные атомы.

The sentence fragment contains two prepositional constructions. "Для разделения молекулы" stands after an adjective, consequently it has to be decided whether it can be the adverb of the given adjective. There are two substantives before the expression "на отдельные атомы" both in genitive case and it has be determined whether it is the adverb of "нужный" or an attribute of one of the substantives. "Нужный" belongs to the 7th group of the adjectives /K.III.=7/, where the serial number of "для" is 1. In the same group no "на" +Acc." occurs. As a consequence the expression "для разделения молекулы" can only be the adverb of the adjective "нужный" . "Разделение" belongs to the loth group of the substantives. Here the serial number of "на"+Acc." is 1. "Молекула" can be found in the 20th group of the substantives. The serial number of "на "Acc." amounts to lo. As a consequence the expression "на отдельные атомы" is the attribute of the substantive "разделение".

b./ Аналогично мы можем говорить об энергии связи молекул в веществе.

The expression "об энергии"stands in the above sentence after the verb "говорить" and depends also on the verb. Further analysis is needed for the expression "в веществе" since there is a triple genitive relation before it and so it has to be decided whether it depends on one of the three substantives or on the verb "говорить".
"Говорить" belongs to the 9th group of the verbs. The serial number of "B+Pr." amounts to 7 in this group. "Энергия" occurs in the 18th group of the substantives and the serial number of "B+Pr." amounts there to 6.
"Связь" belongs to the 35th group of the substantives. The serial number of "B+Pr." is 8 in this group.
" Молекула " is a member of the 20th group of the substantives and the serial number of "B+Pr." is 4 in this group.

Among the groups that may come into account the prepositional expression "B+Pr." has the lowest serial number in the 20th group of the substantives containing the morpheme "молекула" . Consequently the expression " в веществе" is the attribute of the substantive

c./ ... что в природе все заряды по абсолютной величине кратны заряду электрона.

The expression "в природе" stands after a conjunction and is therefore the adverb of the predicate /кратный /. "по абсолютной величине is preceded by a substantive and the question to be decided is whether it is the attribute of the substantive or the adverb of the adjective " кратный " which is the predicate of the sentence. "Заряд" belongs to the 20th group of the substantives. In this group " по" does not occur. We find, however, the preposition "по" in the first place in the 16th group of the adjectives containing "кратный" It is clear that the expression " по абсолютной величине" is the adverb of the adjective "кратный".

III. The result of our examinations

1. There are 190 examples altogether in the Introduction and the First Chapter in which the prepositional construction standing after the substantive depends on the preceding substantive. The analysis carried out on the basis of the verbal, substantival and adjectival groups compiled from the given text the attributive role of these prepositional constructions could be established in 171 cases, while incorrect or questionable decisions were obtained in 19 cases /10%/. When the same examples were analysed on the basis of the groups consisting of the corresponding morphemes in the Introduction + First Chapter + Second Chapter the result was correct in 180 cases and the incorrect or questionable decisions dwindled to 10 cases /5,26%/.

2. Let us now consider some examples for which the syntactic role could be determined only by the aid of the groups set up on the basis of the enlarged corpus.

a./ Первые сомнения в этом появились после открытия..."

According to the grouping of the morphemes of the introduction and the First Chapter /henceforth A./ "B+Pr." stands both in the group of "сомнение " and of "появляться" in the first place. The role of the expression "в этом " cannot be determined on the basis of such a grouping.

According to the grouping, however, resulting from classifying the morphemes of the Introduction, the First and the Second Chapters /henceforth B./ the word "сомнение" belongs to the 2nd group of the substantives where the serial number of "В +Pr." is 1. " Появляться" is the member of the 22nd verbal group where "в +Pr." stands on the fifth place. Now it is clear that the expression "в этом" is the attribute of " сомнение ".

b./ ... будет определяться ... при движении по инерции."

A. In the group of "определяться" the preposition stands in the 2nd, in the group of " движение " in the 3rd place. The analysis yields an incorrect result: the expression "по инерции" is considered as the adverb of the verb.

B. In the group of "определяться" /19th group/ the preposition "по". stands in the 7th, in the group of "движение" /35th group/ in the 5th place. We can now obtain the correct result: the expression "по инерции " is the attribute of "движение".

c./ ... произошел также громадный сдвиг в понимании того ...

A. " в +Pr." stands both in the group of "происходить " and in the group of " сдвиг " in the first place: it is impossible to decide about the syntactic role of the expression.

B. " в +Pr." stands in the group of " происходить " /22nd group/ in the 5th place and in the group of " сдвиг " /20th group/ in the 4th place. A correct decision is now achieved: the expression " в понимании " is the attribute of "сдвиг".

3. It is worth considering some further examples where even the greater corpus /B./ failed to lead to correct results.

a./ Полная энергия, связывающая элементарные частицы в веществе ...

"Связывать" is to be found in the 40th group of the verbs. Here " в +Pr." stands in the third place.

The serial number of в +Pr." amounts to 6 in the group of "частица". As a result we obtain - contrary to fact - that "в веществе " is the adverb of the verb, i.e. we obtain the following translation "connecting the particles in the material" /az elemi részecskéket az anyagban összekötő .../ instead of "... az anyagban lévő elemi részecskéket összekötő .../. In the present case this deviation does not distort the sense essentially.

b./ ...., если электрон в атоме находится в возбужденном состоянии,....

"Электрон" is the member of the 20th substantival group. "в +Pr." stands here in the 4th place. In the group of "находиться " /18th group/ the serial number of "в +Pr." is 2. We obtain as a result of the analysis that the expression "в атоме " is the adverb of the verb "находиться " although it is in fact the attribute of " электрон " We obtain the following translation: "... ha az elektron az atomban gerjesztett állapotban van ..." /if the electron is in an induced state in the atom/ instead of "... ha az atomban lévő elektron gerjesztett állapotban van ..." /if the electron in the atom is in an induced state/. The incorrect translation does not distort the sense here either: "az atomban" /in the atom/ may be considered in the Hungarian sentence as a postponed adverbial attribute.

Similar distortions were found in the remaining 8 cases too, i.e. the incorrect analysis did not distort the sense. This does not mean, however, that in other cases such an analysis would not result in serious distortions of the sense.

4. Since the described system takes into consideration only the required postpositions it cannot be applied in cases when the syntactic role of the prepositional construction can only be defined on the basis of the meaning of the substantive in this construction. E.g. "... они установили печь на новом заводе." /adverb/; "... они установили печь на твердом топливе"/attribute/. To analyze similar cases semantic information is needed on the basis of which even the possibilities of morpheme connections can be predicted. Such a semantic system, however, is not at our disposal as yet. It should be stressed that in the examined text we did not encounter similar examples.

5. Our data show clearly to what extent the correct analysis depends on the length of the examined text. /In our case: if we set up our groups on the basis of a text that is 50% longer, then the result obtainable becomes 5% better. Of course, further increase of the text will not result in such a rapid improvement and we may not expect a 100% result with the aid of this method but if we take a long enough text, we can, with good reason, count upon a result that approximates fairly well the 100% mark. An entirely reliable result can be obtained only if semantic aspects are taken into consideration. At present we lack such a system and so, at least for the time being, the method described above seems applicable even if it has assailable points.

IV. Some further examples

Lastly we give now a few more complicated examples to show the applicability of our method where the prepositional constructions standing after the substantive /or substantives/ play sometimes an attributive, sometimes an adverbial role.

1. ..., теряя энергию при столкновениях с ядрами вещества и на рождение частиц.

The example contains three prepositional constructions.

a./ "при столкновениях" which may be the adverb of "терять" or the attributive of " энергия ". In the group of " терять " /37th group/ the preposition "при" stands in the 4th place and in the group of "энергия" /18th group/ in the first place. The expression is consequently the adverb of the verb "терять".

b./ "с ядрами" which may be the adverb of the verb "терять" or the attribute of "столкновение" . The preposition "с" stands in the 10th place in the group of "терять" /37th group/ and in the first place in the group of "столкновение" /16th group/. The expression quoted is therefore the attribute of " столкновение".

c./ "на рождение" which may be the adverb of "терять" or the attribute of " вещество ". The preposition "на +Acc" occurs in the 9th place in the group of "терять" and in the 10th place in the group of "вещество" /20th group/. The expression is consequently the adverb of "терять".

2./ Тогда из соотношения между энергией и импульсом для случая частицы без массы покоя ... заключаем....+

a./ "из соотношения" stands after a conjunction, it is consequently the adverb of "заключать".

b./ "между энергией и импульсом." "Соотношение": 21./8/; " заключать " 25./-/. The expression is the attribute of "соотношение"

c./ "для случая". "Энергия" 18./10/; " импульс " 18./10/; "заключать" 25./9/. The expression is consequently the adverb of

d./ " без массы " The morpheme " случай " did not occur in the

---

+ In the following we put the number of the group and,in parantheses, the serial number of the corresponding preposition - for the sake of brevity - after the morpheme /the complement of which the given prepositional construction can be/. The prepositional construction is always the complement of the morpheme that is followed by the smallest number in parantheses.

examined text with prepositional requirement but in virtue of its meaning it is unlikely that the prepositional attribute "без" is more frequent after it as after "частица " . "Частица" : 16./6./; "Заключать" 25./-/. The result is that the expression is the attribute of " частица ". A larger corpus would also clarify the governments of " случай ".

3./ По закону сохранения импульса после превращения импульс частицы Б должен быть равен сумме импульсов обеих частиц до превращения.

a./ " по закону " stands at the beginning of the sentence, it is the adverb of "равный".

b./ " после превращения ". " закон " : 2I./-/; "сохранение": 35./-/, "импульс": I8./9/, "равный": I7./7./. The expression is the adverb of " равный ".

c./ " до превравщения ". " сумма ": I9./-/; " импульс " : I8/5/; " частица " : I6/-/; " равный " : I7./IO./. The expression is the attribute of " импульс ".

This example is interesting because "после превращения" as well can be considered as the attribute of "импульс" because it stands before the predicate: in the second part of the sentence one might be tempted by the analogy of " импульс до превращения" The mechanical analysis however revelas clearly that it is the adverb of "равный".

4./ ... проекция спина на направление магнитного поля в первом случае равна ...

a./ " на направление " . " проекция " : IO.I./; " спин " : 36./-/; " равный " : I7./-/. The expression is the attribute of " проекция ".

b./ "в первом случае" . " направление" : 34./-/; " поле " : 8./-/; " равный " : I7./I./. The expression is the adverb of "равный".

5./ ... что дальнейшее продвижение по ступеням элементарности будет напоминать переход от молекулярной ступени к атомной или переход от атомной ступени элементарности к современной.

a./ " по ступеням " . " продвижение " : I9./8/, " напоминать " : 9./9/. The expression is the attribute of " продвижение " .

b./ " от молекулярной ступени " " переход " : 25./7/: " напоминать " : 9./-/.

The expression is the attribute of "переход".

c./ " к атомной " . " переход " : 25./4./;
"напоминать": 9./-/. The expression is the attribute of "переход".

The analysis of the second part of the sentence may be carried out analogously to b./ and c./.

V. The data obtained from a rather restricted corpus did not make it possible to check the method with respect to other texts. It is clear, however, that the correctness of the method can only be proved if it can be extended to texts that are different from the corpus /or corpora/ out of which the data are obtained. Such an extension remains to be done, therefore we postpone further refinements of the elaborated method to a later paper.

## References

1 Sipőczy Győző: Az elöljárós kifejezések programozási kérdései. Általános Nyelvészeti Tanulmányok. II. Akadémiai Kiadó, 1964. 231-242 pp.

2 Hell, G.: Recognition of Nominal Groups in MT. from Russian into Hungarian. Computational Linguistics, I. 5-1o7 pp.

3 Л.Н. Иорданская: Два оператора обработки словосочетаний с "сильным управлением". "Москва, I96I."

4 E. Daum, W. Schenk: Die Russischen Verben. Leipzig, 1954.

# Computer simulation of human behaviour and concept formation in music composition

by

E.N.Ferentzy

Summary

The way of attack referred to in the title is to register evidence on human behaviour during music composition on protocols and trying to produce models apt to yield similar protocols if run on a computer.

Of such models one that is based on tonal equivalence is discussed in more detail.

As a general principle for choosing from possible equivalence classifications the simplicity test is found. That is, we measure the simplicity of the series which is got if equivalence classes are substituted for corresponding primary elements.

Ranking by simplicity is realised by comparing transition probability matrices or phrase structure language representations.

The second part of the paper deals with the recognition of chord patterns in figurative formations. Experiments pointed to the necessity of introducing the notion of concept formation by spheric extrapolation. The place of this procedure in the framework of concept learning theory is discussed.

X

Formal analysis of composed music has been carried out for a rather long time. /See literature./ Nevertheless, this analysis consists more in a formal **mat**hematical treatment of the composed work, partly to detect laws, partly to get enough information to be able to simulate those compositions.

In the present paper a different way of attack is reported. In this method the general attitude is very similar to that which has led to the development of the General Problem Solver of Newell, Shaw and

Simon [4]. Individuals who are given elementary music composition problems to solve are questioned, their answers are registered on protocols, a formal model is built on certain hypotheses, and a protocol got from the actual computer running of a programmed version of the model is compared with the protocol of the individuals. It is a very advantageous situation that in investigating human behaviour in elementary music composition we are not restricted to experiments especially designed, but we can refer to the whole literature on the teaching of traditional harmony. These books not only contain laws concerning rules of composition of classical music, but very often go into the detailed description of how these rules would be applied in different situations. Such descriptions can be with right regarded as protocols of the verbalised thinking process in elementary music composition.

To put it more precisely what under elementary composition here is understood: we mean by that the different tasks given on courses for students in music composition, so e.g. to construct a four-voice accompaniment to a given melody, or to compose the melody beginning on a given bass line.

Although research in this field has not yet yielded results which could be regarded as a complete description of human behaviour of elementary music composition, some aspects of it can already be discussed.

Tonality as a unifying concept of the understanding of classical music /we are not concerned in the present research with music that cannot be investigated in the general framework of classical music/ was and is accepted without discussion. So it is not surprising that in the search for underlying laws of musical composition of elementary type with formal methods this concept has been found to have an overwhelming importance.

If we regard the tonal theory of the different modes and of the Subdominant, Dominant and Tonica functions, this can be regarded formally as the introduction of certain equivalence relations among the different chords, so I, and partly VI and III are regarded as T; V, VII, and partly III are regarded as D, IV, II, and partly VI are regarded as S. These equivalence relations unify even larger sets of the various possible realizations in actual voices of the chords into one equivalence category. The introduction of these equivalence relations into the interpretation of the music score results in the establishment of a very simple pattern of the series of T, D, and S functions, which are got if the chords of the music score are replaced by their function equivalents. In fact the formation of tonal feeling can be explained by the fact that relations of possible substitutions among different chords have been developed by composers and trained into the ears of listeners for the purpose to ensure a simple basic pattern of music which can serve as a basis of variations. Because if if is true that several chords and even a larger number of actual realizations of these chords evoke the same tonal impression, then during the perception of a music piece listeners will have the feeling of a cyclic movement of a rather simple structure and this will ensure that orientation in the meaning of the chord actually heared is simply achieved. So we might suppose that the existence of some type of functions, of some type of equivalence relations among different constituents of music is an obligatory requirement for the purpose that music should be understandable. This statement is confirmed by Schoenberg also for twelve-tone music /see in [3] / naturally with the addition of the fact that these functional relations of twelve-tone music are quite different from those of classical music, and today they are yet unknown and will be known only after analysis has been carried out on a large set of twelve-tone music pieces.

From a formal point of view the importance is in that finding that the formation of equivalence relations among constituents of music must point to a simplification of the structure of the music pieces, for which these relations are introduced. Here we simply mean that transcribing the constituents of music, say the chords, by their function equivalents, the resulting series must have a simpler structure than the original. This fact is important because the simplification of a formally given stimulus series if transcribed by the application of certain rules into another series can be easily measured by mathematical means. So without any a priori information concerning the equivalence relations a computer can be programmed to introduce such relations for the transcription of an input stimulus series and then to test whether this leads to a simplification in the structure of the series. If no simplification is achieved, the equivalence relation experimented with is rejected and another is generated. This process is continued until a feasible equivalence relation is found. Mathematically this procedure is equivalent to "clumping" of transition frequency matrices, and generally the theory and results of the decomposition of stochastic automata are employed. /See [8] /.

If there is a unique equivalence relation yielding the structural simplification then this can be accpeted as the most suitable functional interpretation of the constituents of the music in the style that has been analysed. If there is no unicity of feasible equivalence relations then it can be concluded that in the functional interpretation other factors play a part too, different from the tendency to end with a simple structure.

Simulation experiments are carried out on the computer National-Elliott 803 B. A detailed report of this work is left to a forthcoming paper but it should be noted that even the results obtained so far from the analysis of chord series of harmony course books make it possible to justify also from a formal point of view the classification of the chords I-VII into the traditional equivalence relations T, D and S. Since, if transition probability matrices are formed of chord series, after the introduction of this equivalence relations the series are transformed into the consecution of the three types of cadences DT, ST, SDT, that is they show a very simple structure. According to this the "clumping" of the transition probability matrices of the chord series results in a transition probability matrix of the S, D, T functions of the following simple type:

| | T | D | S |
|---|---|---|---|
| T | X | + | + |
| D | + | X | O |
| S | + | + | X |

Fig. 1.

In the figure + represents positive transition probability, X has the meaning that we regard e.g. T after T as a prolongation of the function T, and so a consecution T after T is not defined.

There is another approach which in this simple case has led to the same result, but which has the greatest importance in music analysis. We think here of the cyclic representation method of the grammars of phrase structure languages introduced by Solomonoff in 1959 /see [4] /, the basic idea of which being developed earlier by Minsky.

Repetition being the most important feature of music, as stated unanimously by music theory, a representation in "cycles", and "cycles upon cycles" seems to fit excellently the nature of music.

If phrase structure grammars are introduced at random /as equivalence relations above/, that one will be chosen which allows the simplest interpretation of the original musical sentences if expressed in that grammar. This could speed up the development of the "teacherless" learning of grammars.

What can formal music analysis add to concept learning theory?

Concept formation problems are encountered necessarily if someone is interested in computer simulation of human behaviour in music composition. This is true even in the simple case when we want to develop methods able to transform a real music score of music literature into a series of chords which is a prerequisite of the above analysis. Perhaps the most obvious first impression is when somebody is looking formally on a real music score that nothing similar can be found to what is called a chord in a harmony textbook. Instead of that we find rhythmically and melodically figurated formations, so that in the end each bar differs from the others in some minute detail. So the need arises to develop formal methods which are able to recognise chords or other specified constituents of music if they are hidden by the figurative richness of the actual music score.

Now if we compare several figurative formations with the purpose of finding their common character, we are already dealing with a concept learning problem.

Stimuli regarded in concept learning are here represented by elements acting in temporal consecution. So notes with different characteristics are communicated to the listener ordered in time.

If the actual problem is to learn the concept of a certain chord, this concept has to be abstracted from different stimulus series, which are the figurative representatives of the chord.

The analysis of figurative chord formations regarded as stimulus series makes possible the establishment of consequences which refer to the structure of concept learning models in general.

The formation of super-stimuli and the theory of concept learning.

Firstly the standard concept learning situation as generally understood is formulated.

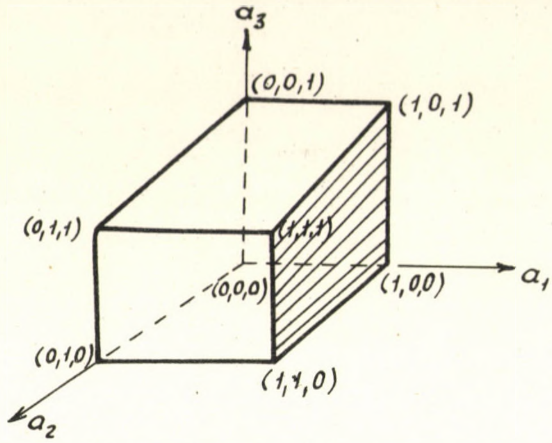
In concept learning situations it is always supposed that different perspicuous aspects of the object analysed are acting as distinguishable stimuli. If $n$ aspects are taken into account, then the object might be characterised by ab $a_1, a_2, \ldots, a_n$ vector, where $a_i$ represents the value of the i-th stimulus. For the sake of simplifying the discussion let us suppose that the stimuli might have only two alternative values, which might be denoted by 0,1.

In this case the universe $E^n$ is the $2^n$ element set of all possible $a_1, a_2, \ldots, a_n$ vectors. Positive and negative instances used to teach a concept are taken from this universe. To further simplify the discussion we will assume that the concept to be learned is represented only by positive instances.

The standard concept learning approach presupposes that the concept to be learned is in the most complicated case a disjunctive concept. A disjunctive concept can be regarded as the union of several conjunctive concepts not necessarily exclusive.

Now it is very easy to see that using the terms of linear algebra, a conjunctive concept is nothing else than a subspace of the $2^n$ element universe. Let us demonstrate this on an example. Let $n = 3$. This has the advantage that the whole universe can be represented in the three-dimensional Euclidan space. In fact the whole universe consists only of the 8 lattice points in the positive ortant having exclusively 0 or 1 coordinates. The elements of the universe are generated if we allow for the stimuli $a_i$ in the $a_1\ a_2\ a_3$ stimulus vector to take on the values 0,1 independently of each other. Now, a conjunctive concept can consist in the rule that for all positive instances of the concept $a_1 = 1$ obligatorily, while $a_2$ and $a_3$ take on values in all possible combinations.

Let us look at the figure below. It is clear that the conjunctive concept represents a face of the cube, which is the representative of the universe. /Naturally only the integer points of the cube are constituents of the universe/. If we fixed two $a_i$ coordinates, the corresponding concept would be reduced to a subspace of a lower order, it would be represented by an axe, that is by a one-dimensional subspace of the universe.




The case of n = 3 was perhaps well suited to demonstrate how concepts are represented by subspaces, but for the concept learning task an 8 elements universe is too trivial. Therefore let us suppose that n is fixed as a larger number, and we are confronted with the task of learning a conjunctive concept on the basis of given positive instances.

More precisely the situation is the following: the concept to be learned is known to be conjunctive concept, that is it consists of all the points of a subspace of the universe. In psychological concept learning experiments this is ensured by telling the subject that he has to learn a conjunctive concept. No, in order that the task should not become trivial, it is necessary that the concept to be learned should not be given by all the elements of the corresponding subspace. So in this case the concept learning process is the following.

The existence of a concept-subspace equivalence is presupposed. Positive instances therefore are regarded as samples from this subspace. The only problem is to find a subspace which is most compatible with the positive instances. In other words there is an extension or extrapolation done with the help of the positive instances. This extrapolation - if the basic concept-subspace equivalence is once accepted - does not cause any serious problem.

The same situation arises when we regard positive instances as representatives of a disjunctive concept, that is, as taken from a union of subspaces of the universe. It should be noted that the subspaces in question might have common elements. If this basic assumption is made, the concept learning problem merely consists in finding such a union of subspaces as fit with the highest probability the set of positive instances. It is clear that such a fitting of subspaces and the selection of the most probable one can be simply done by a human subject or can be simulated by a computer.

There is one thing we should not forget in applying the above techniques. The whole procedure described above is based on the assumption that the concept is represented by the union of subspaces of the $2^n$ element universe of the possible values of the immediately given stimulus vector. Moreover, extrapolation on the basis of observed positive instances is guided by the principle that a disjunctive concept cannot consist of a large number of conjunctive members, that is, in the union of subspaces only a few number of subspaces take part.

Now this procedure of defining the concept as a union of subspaces of the universe of possible stimulus vectors clearly represents one way of extrapolation with the help of the positive instances of the concept. If we regard the problem as a purely formal task the limitations of this way of generalization become quite obvious. It is by far not sure that points in an Euclidian space can be embedded in an optimal way in a union of subspaces.

A large number of experiments carried out with stimulus series taken from the field of music has confirmed that the best way to generalize positive instances is definitely not to embed them in a union of subspaces. The embedding process - as this is most obvious in the case of conjunctive i-th stimulus there are two possible cases. The one case is that the value of the i-th stimulus belongs to the essence of the concept and so it must be the same in all positive instances. The second case is that the value of the i-th stimulus does not belong to the essence of the concept and so it must be oscillating between the values 0 and 1. In the case of disjunctive concepts, which must consist of only a few conjunctive members, similar requirements can be found. Musical stimulus series have been analysed according to this technique. Analyses have led to a complete failure, simply because there could be found no stimuli in the immediately perceived stimulus series which would remain constant. The general picture is that the whole series is fluctuating.

Therefore other ways of extrapolation on the set of positive instances had to be thought of. It was not too difficult to find a general way of attack that could be employed. It was found that although all the stimulus values are fluctuating, if e.g. the number of 1-s is counted for every positive instance, this number behaves rather regularly. Starting from this experience, a method of extrapolation could be established which has led to a new way of concept generalisation. Concepts established by this new method could be called by the way of their formation "spheric concepts".

The basic assumption in the extrapolation procedure leading to spheric concepts is that the most simple concepts of this kind, corresponding to the conjunctive concepts discussed above are characterised by having all their instances on a sphere. More precisely, on the part of the sphere in the positive ortant. Let us look at one example for such a spheric concept. Given the $2^n$ element universe of $a_1, a_2, \dots, a_n$ stimulus vectors, let us define a spheric concept in the following way:

$$a_1^2 + \dots + a_n^2 = 1 .$$

In other words the concept consists of all the stimulus vectors which must have a 1 in a single position and 0-s in all the others. Where the 1 element is placed, this is exactly the varying parameter constituting the concept. Naturally this concept having such a simple representation in the spheric form, could also be expressed as disjunctive concept, but in a rather complicated way. Now if among the positive instances we could find a large number of vectors having 1 in a single position, and 0-s in all the other positions, and the single 1 occurred for different $a_i$ stimuli in the $a_1, a_2, \dots, a_n$ stimulus vector, it would be quite natural to use the spheric extrapolation technique instead of trying to extrapolate with the help of the union of subspaces.

As disjunctive concepts consist of several subspaces, so spheric concepts may consist of the union of several spheres. The important thing is that this extrapolation technique is also justified only as long as it can yield in the end concepts consisting only of the union of a few number of spheres. This restriction arises from the fact that our aim is to simulate human concept learning. The human brain always finds the optimal way to memorise concepts with minimal effort. Human intelligence is surely not restricted to carry out extrapolation on the basis of immediately given stimuli according to a subspace system as carried out in disjunctive concepts. It is also sure that extrapolation on the basis of immediately given stimuli according to a subspace system as it is supposed if we think of human concept formation as carried out in disjunctive concepts. It is also sure that extrapolation is not done in every case according to the spheric metrics introduced here. Nevertheless the large number of experiments concerning concept formation in music learning proved that fixed stimulus values of fixed stimuli play almost no role. We mostly react on the fact that in the consecution of stimuli a certain stimulus value has concurred or not. To find examples of this, we need not go as far as music. It is enough if we regard the dynamic factor in everyday speech. In this respect speech consists of spelling units. Each spelling unit is essentially characterized by one accented syllable around which un-accented syllables are grouped. How this grouping is arranged, that is the serial number of the accented syllable in the spelling unit, is dependent on the particular situation. Therefore, if we do not regard the meaning, the melody etc. of the spelling unit, but only the dynamic factor, so we must constatate that the general concept of a spelling unit from the point of view of accent behaviour is a spheric concept.

It is evident that if spheric concepts are thought of as composed of the set of elementary spheric concept /an elementary spheric concept is defined by all the vectors, having i stimuli with value 1 and j stimuli with value 0, $i + j \quad n$/, then they can be regarded as usual disjunctive concepts. But in this context the smallest building stone is an elementary spheric concept, and not the immediately given stimuli. Our present subject here is the study of extrapolation based on the immediately given stimuli.

For the study of concept learning based on immediately given stimulus series two procedures were discussed in this paper. The first is the traditional disjunctive concept approach extrapolating in union of subspaces of the universe. The second, introduced here, extrapolates according to a spheric metrics, leading to a concept representation by a union of spheres of the universe. The basic assumption in both cases is the assumption of an underlying metrics. For adequate characterization of music phenomena are both unsatisfactory. There is not much hope either that the straightforward introduction of other type of metrics will solve the problems. This situation is regarded to be a lucky one insofar that research is necessarily pushed toward the investigation of the formation of metrics in general.

<u>The process of the metrics-formation in general.</u>

To see there clearly we have to set the basic concept learning task precisely.

Learning a concept, we are deciding that a yet unknown point of"$E^n$ does belong to the concept A, which is a subset of $E^n$".

According to this, if we decide that a point x belongs to A, then we put only in other words that we found the point x to be an element of a set A', which has the greatest probability to be identical with A.

How we derive this information on the probability of a set of $E^n$ for being identical with A?

To give a final answer to this question would be early today, but two possible investigation-lines can be proposed.

1./ The strategy is based on the set positive /and negative/ instances given for concept A.

Positive instances must exhibit common features, characteristic on A. Therefore, if a transformation is found mapping the positive instance /and only those/ into each other, so this transformation must leave characteristic features of A intact.

If an unknown point can be derived with probability P from the given instances by the $\Gamma$ transformation, then it is itself with probability P on element of A.

Using this method, we have generated the set A', which consists according to our opinion, of points of A.

Let us denote the sample of positive instances by S, let denote the derived transformation, so A' can be written formally in the following way:

$$\Gamma X = \left\{ y \mid y = \Gamma x,\ x \in X \right\}$$

$$A' = S \cup \Gamma S \cup \Gamma^2 S \quad \ldots = \sum_{n=o}^{\infty} \Gamma^n S$$

2./ The strategy uses also information different from that contained in the sample of positive instances for concept A.

In most cases the sample of instances for a concept A is not "rich" enough either to allow unambiguous conclusions on the elements of A.

The plus information we use is that the concept to be learned is element of a group of concepts, the elements of which are known to us.

The elements of a group of concepts can be given e.g. by a generatic set of basic concepts.

The above strategy is applied in all concept learning experiments, when the subject searches the concept as a disjunctive concept, that is as point set consisting of union of subspaces.

It is important to remark that the introduction of concept - group strategy does not solve the problem, only transfers it to a higher level.

We have seen that groups /or types/ of concepts introduced in a straightforward way /disjunctive, spheric concepts/ are not satisfactory for extrapolation in every situation.

Therefore we have to answer the question how human brain acquires knowledge on the structure of the type of concept actually valid.

The first phase of the process of learning a type of concept consists clearly in showing explicitly to the subject several concepts belonging to the type.

Here we are again at the central problem of concept learning: on the basis of the explicitly given concepts we have to conclude on the remaining elements of the type of concepts.

One fact seems to be obvious: if some point sets are known as the elements of the concept type, then union and difference of those sets are also elements of the concept type.

The great advance also here /as in the case of learning a single concept/ can be hoped of the determination of the transformation in general, suitable to map the known elements of the concept type into each other.

Let us look on the following example:

$$n = 10, \quad E^n = 2^{10}$$

The subject has been shown the following concepts:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| x | 1 | x | x | 1 | 1 | 0 | 1 | x | x |
| 1 | 0 | 1 | 0 | x | x | 1 | x | x | x |
| x | x | x | x | x | 1 | 1 | 1 | 1 | 1 |
| x | x | 0 | 0 | 0 | 1 | 0 | x | x | x |
| 1 | 0 | x | 0 | 0 | x | 1 | x | x | x |

x = free bit

The subject observes that for the transformation "$\sigma$= permutation and negation of the fixed bits", the elements of the concept-type are invariant. So it concludes that e.g. the following concept belongs to the concept type as well:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| x | 1 | 1 | x | x | x | x | 0 | 1 | 0 |

This way the general notion of "conjunctive concept" is formed.

## L i t e r a t u r e

[1] Pinkerton, R.C.: Information Theory and Melody /Sci. American, 194/2/:77, 1956/

[2] L.A.Hiller, Jr. and L.M.Isaacson: Experimental music Mc Graw Hill Co. 1959./

[3] A.Schoenberg: Die formbildenden Tendenzen der Harmonie /B. Schott's Söhne, 1957./

[4] A.Newell, J.C. Shaw and H.A.Simon: Report on a general problem-solving program
R.Solomonoff: A new method for discovering the grammars of phrase structure languages
Information processing, edited by UNESCO, R. Oldenburg, München, 1960.

[5] Earl B. Hunt: Concept learning /John Wiley, 1962./

[6] Computers and thought, edited by E.A.Feigenbaum and J. Feldman /Mc Graw Hill, 1963./

[7] A.Newell: Learning, generality and problem solving
C.Vossler and L.Uhr: Computer simulations of a perceptual learning model
Information processing 62, edited by M.Popplewell /North Holland Publ. Co. 1963./

[8] Glenn C.Bacon: The decomposition of stochastic automata. Information and control. 1964 Sept. /Ac. Press, New York/

[9] E.N.Ferentzy and M.Havass: Human movement analysis by computer. Electronic choreography and music composition. /Computational Linguistics III. Budapest, 1964. Computing Centre of the Hungarian Academy of Sciences./

[10] M. Havass: A simulation of the music composition. Electronically composed folk-music. /Computational Linguistics III. Budapest, 1964. Computing Centre of the Hungarian Academy of Sciences./

[11] E.N.Ferentzy: On formal music analysis-synthesis: its application in music education. /Computational Linguistics IV. Budapest, 1965. Computing Centre of the Hungarian Academy of Sciences./

# On formal music analysis-synthesis: Its application in music education

by

E.N. Ferentzy

## Summary

The paper presents a comparative survey of formal methods applied in music research. The method of transition probabilities is found to play a distinctive role.

Formal methods are developed as to take into account more differentiating viewpoints coming from musicology. The method and computer program "analysis in rhythm-type" are worked out.

The theory is applied to a practical investigation of multi-voice pieces for children choirs, so results have immediate realisation in the better knowledge of exactly that type of music which is lacking today for the youth in the age of changing voice.

The method is tested on outputs obtained both in mathematical form as probability matrices and in music form as computer composed two-voice children choir songs. /Music-scores presented./

Tests support theoretical considerations and show the way of further developments.

## 1. Introduction

This paper will present a comparative survey, the development and applications of mathematical algorithms used in music analysis.

The mathematical algorithms to be presented here are based on the research and systematising methods of musicology.

The real revolutionary change is due to the fact that even traditional analysis is applied having in mind the availability of computers. These computers have, from the point of view of the tasks to be solved in this case, an infinite memory and infinitely rapid operation time, so that in planning research it need not be considered what computational labour is to be expanded. Investigations which could not be planned in the past can be carried out today without any difficulty.

On the one hand, the result of the investigations is, as in traditional research, general laws. On the other hand, computers make it possible to present in an explicit way all the variety of music ideas which can be found in a whole set of music pieces.

It is important to set up such a catalogue of music ideas and basic motives especially in those branches of music composition where the conditions are still unknown, and which fields of music composition have to be developed. To the branches to be developed belongs the very important field of music for the youth in the age of changing voice.

As has been shown in [1], in the age of changing voice, especially in the case of boys, the range of the pitch in which a boy can sing is very limited. Moreover, boys can sing only in one of the three ranges soprano, bass and cambiata, and these ranges have not a single note in common. One of the ranges does not comprise more than 4 notes.

Therefore, it would be very important to have three voice music, where the single voices are within the limited ranges pointed out in [1].

The fact is that at present there are not enough composed music pieces to meet these requirements. Therefore, it would be very interesting to investigate what type of music ideas and basic motives are employed by existing music pieces fulfilling the requirements which are necessary to present a music for boys in the age of changing voice. From such a catalogue of basic motives it would be possible to find the ways in which this field of music could be developed.

The paper regards it as its purpose to give as good demonstration of the mathematical formalism used, as possible.

A music score of a computer-composed two-voice children choir song is presented in the App.

It should be noted that at the present stage of research it was not our aim to formalize larger music units, but the main concern has been to find the basic music ideas, and the ways in which these elementary units are connected with each other. The computer composition given in the Appendix reflects exactly how the original music would look like if the larger compositional ideas were left out and only the elementary laws of aesthetics would form the basis of composition. To investigate the larger compositional units and to formalize the laws or tendencies governing them are the main concern of the present research of the author and will be reported in another publication.

## 2. Comparative presentation of the method of transition probablities.

### 2.1. The transition probability matrix for the mathematical representation of music information.

For the mathematical representation of music information the transition probability matrices have a distinguished role.

Before attempting to prove this, we have the following remark to make. The transition probabilities as well as the statistical parameter belong to the concepts of mathematical statistics, and so they are quite abstract. Therefore nothing can be said in general concerning the value of their application. There is such an evaluation in every concrete instance.

What we assert is that the probability matrix is an excellent means for the representation of music information. Only in this respect can it be said that the other tools of mathematical statistics are of secondary importance.

The insufficiency of the other methods /first of all in the methods applying statistical parameters directly on coded music, such as mathematical means, dispersion, kurtosis etc./ is that they try to express with one single data the manifolded nature of music. This can be justified in a preliminary investigation, but in every case it must be regarded as no more than an approximative method at best.

The method of the transition frequencies, on the other hand, works on the principle of sorting /see later/. In this way the method will differentiate the information, and make explicit very important rules and interdependencies which have been implicit in the music analysed.

The real purpose of the mathematical representation is that it should become possible later to carry out mathematical-logical operations on the information stored. Accordingly, our task is to find a final mathematical representation that is characteristic of the dependencies or the diversity of those pieces.

Now, if the original representation has been given in a very contracted form /e.g. as a simple statistical parameter of the pitch etc./ these assertions concerning the regularities valid for the set of pieces analysed, have very questionable value.

On the contrary, as the analysis is done in a more and more sophisticated way /there is a transition probabability matrix with many elements produced, or a whole set of transition probability matrices/, the better we may trust that the summation of this individual matrices will lead to a result representation of which it may be said that it expresses the common features of the pieces analysed.

To evaluate these common features on the basis of the mathematical representation of the analysed information, the statistical parameters can justifiably be applied. /See the introductory remark on the general character and practical application of formal mathematical tools./

The value of the representation of music information with transition probability matrices and the value of the operations with these matrices can be judged also by applying on the result matrices the method of electronic composition with computer programmed random number generator. This possibility is not a negligible one.

Can a better form be imagined for the evaluation of the results of an analysis carried out with a formal mathematical method, for musicians, than to activize the music information analysed and to present it in a music form?

What was hoped to be attainable with the help of the analysis? We wanted to find common musical features of a group of musical works, or we intended to find connecting links between two types of music.

If the analysis and the computer representation of music have been correct then in the first case the music got in the end has to reflect the basic common features of the music pieces analysed; in the second case it has to reflect the properties of the transition type music between the two styles investigated.

Mathematical proof of the statement on the rich information content of the representation with transition probability matrices of music information.

The aim specified in the title is attainable if on the basis of the generally defined transition probability matrix, the probability distributions and statistical parameters used in the representation of music information can be deduced, and in the deduction summation occurs. This last mentioned fact clearly shows that going over to representations other than that with transition probability matrices causes a contraction of information. The deductions which follow will prove that this is so.

The following deductions naturally do not mean that this is not only a possible but also an obligatory way to determine those distributions and parameters. The distributions and parameters can be calculated directly on the basis of the original music information too. It should be remarked that by music information in general we simply mean a coded music score or a digital transcription of analogue registered tune-diagrams.

The deduction of distributions other than the set of distributions forming the transition probability matrix. The frequency distribution of basic elements.

/Pitch, duration of notes etc./

The frequency distribution of the occurrence of the basic elements can be very simply got from the transition frequency matrix which is even more simple to get than the transition probability matrix:

$$\left\{P_{ij}\right\} \quad \begin{matrix} i = 1, \,..\, , n \\ j = 1, \,..\, , n \end{matrix} \Longrightarrow \left\{ \sum_{i=1}^{n} P_{ij} \right\} \qquad j=1, \,..\, , n$$

The deduction of the frequency distribution of intervals of consecutive basic elements.

/The pitch or duration of the notes./

This can be done with the help of the transition probability matrix and with the help of the probability distribution of the basic elements: $\left\{p_j\right\}$ , which is already a derived distribution as shown in the foregoing section:

$$\left\{P_{ij}\right\} \quad \begin{matrix} i=1, \,..\, , n \\ j=1, \,..\, , n \end{matrix} , \quad \left\{p_j\right\} \; j=1 \,..\, , n \Longrightarrow \left\{ \sum_{j=\max/1,1+i/}^{\min/n,n+i/} p_{/j-i/j} \cdot p_j \right\}$$

$$i = -/n-1/, \,..\, , 0, \,..\, , n-1$$

The deduction of the value of the statistical parameters.

The mathematical means.

If the basic elements are numbered 1, ....,n, so the mathematical means M can be simply expressed with the probability distribution in the following way:

$$M = \sum_{j=1}^{n} j \cdot p_j$$

The dispersion /standard deviation/.

The elements numbered 1, .... , n the formula of the standard deviation is:

$$\sigma = \sqrt{\sum_{j=1}^{n} / j - \sum k \cdot p_k /^2 \cdot p_j}$$

The kurtosis.

The elements numbered 1, ... , n the formula of the kurtosis is the following:

$$K = \frac{\sum_{j=1}^{n} / j - M /^4 \; p_j}{\sigma^4}$$

Correlations coefficients.

With the help of the transition probability matrix $/p_{ij}/$ and with the help of the frequency distribution $/p_j/$, the one lag correlation coefficient can be expressed as follows:

$$1 = \frac{\sum_{i=1}^{n} \; \sum_{s=1}^{n} i \cdot j \cdot p_{ij} \cdot p_j}{\sum_{j=1}^{n} /j - \sum_{k=1}^{n} k \, p_k /^2 \cdot p_j}$$

To deduce correlation coefficients for greater lags transition probability matrices of greater depth must be used.

This last remark shows most clearly that even if the transition probability matrix is the richest in information content among all the other means of representation that does not deprive in any case the other methods listed of their importance in preparatory information processing. If nothing is known of the regularities of a music piece or a set of music pieces the great expense of computer time which is required to make a multilevel transition probability matrix would not be justified. It is quite natural that in this case the correlation function is first calculated for several lags to see if there are any interdependencies between the notes following each other and how far in depth these interactions go. Only after this correlation analysis has pointed out the number of foregoing notes playing a role in the choice of the investigated note will the transition be registered.

2.2. The method of the transition frequencies is a very descriptive tool.

To visualise the generation of transition frequencies which are the basis of the calculation of the transition probabilities /in fact there is only one division required to go over from the transition frequencies to the transition probabilities/, let us regard a hundred cubes numbered from 1 to 100.

The cubes may have the colour green, blue and red /so as the set of notes in one octave consists of the notes c,d,e,f,g,a,h./

The next task is sorting. We may be interested in several types of sorting and even in the case of one type of sorting different depths might be of interest.

1. What is the number of the red, green, blue cubes in the series?

To answer this question there are three cases prepared, and the cubes are sorted in these cases:

| |
|---|
| R E D |
| G R E E N |
| B L U E |

If only the number is taken of the cubes contained in the different cases, so this set of numbers is called the frequency of the occurrence of the red, green and blue cubes.

What has been demonstrated above is also generally the method of carrying out an analysis of the frequency of occurrence of a certain set of elements ordered in a single row, such as the notes are in a tune.

2. How often do we find a red cube that way that the previous one /in the order of numbering/ is also a red cube, or not a red but a green or a blue cube?

For answering this question the cases prepared previously have to be divided into sub-cases.

| A N C E S T O R S | | |
|---|---|---|
| R E D | G R E E N | B L U E |
| | R E D | |
| | G R E E N | |
| | B L U E | |

If only the numbers are regarded contained in the subcases-system, so this two-dimensional set is called a transition frequency matrix. This is also the general way to come to a transition frequency matrix, if the elements sorted are not cubes but elements of any type, e.g. music identifiers.

3. In the sorting described above the cubes with the same colour and also having an ancestor element with the same colour are put in the same sub-case. If we have the impression, or perhaps a correlation analysis has proved, that the actual colour of an investigated cube is influenced also by the colour of the cube standing with two elements before it in the row, we may prepare a new, more refined system of sub-cases by a new sub-division of the cells of the previous system.

So the content of the previous sub-case defined "red cube with a green 1. ancestor" can be sorted in three new sub-cells; according to the colour of the 2. ancestor.

So a two-level transition frequency matrix can be got. Theoretically the procedure can be continued ad infinitum. So the above procedures are apt to explain the inter-actions in any ordered row in any depth, as the example of the ordered row of multicolour cubes shows.

## 2.3. The representation of the transition frequency matrix with p-graph.

Transition frequencies are determined for so many different kinds of elements /pitch of notes, duration of notes, duration of pairs of consonant notes, series of duration of notes, intervals etc./, that it is necessary to have an exact definition of what transition analysis is. A definition could be given with the help of the abstract mathematical set theory, but it is preferable to use a branch of mathematics which is more descriptive of such nature is the graph theory.

The mathematical graph is simply a set of points which are connected with lines, that is, with arches. A p-graph is such a graph, where it is allowed that from one point to another point in the same direction there should be more than one arch running.

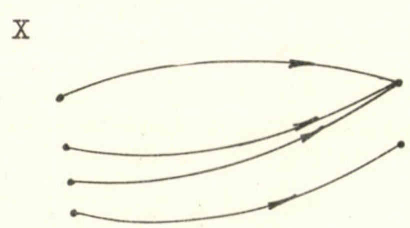
With the help of a $/X,Y\Gamma/$ p-graph the transition frequency matrix can be defined in generality as follows:

X = the set of the identifiers which are regarded as "ancestor" elements in the analysis,

Y = the set of the identifiers which are regarded as "successor" elements in the analysis,

$\Gamma$ = if in the material analysed there are m cases such that after an element x an element y stands, then in the corresponding graph from the point x there are m lines directed to the point y.

The graphical illustration of the above definition:




Concrete examples for the different realisations of the general p-graphs representing transition frequency matrices.

a. One-level transition frequency analysis on a basic set containing n elements.

X = the basic set of elements; 1, .... n.
Y = the basic set of elements; 1, .... n.

b. Two-level transition frequency analysis on a basic set containing n elements.

X = all the possible two-element ordered sets which can be prepared from the basic set; 1, .... $n^2$.
Y = the basic set of elements, 1, .... n .

c. Analysis of the consonant notes /or rhythm pattern of consonant bars/ of two-voice music.

X = the set of pairs of notes, /pairs of rhythm patterns/ the first elements of the pair belonging to the first voice, the second element of the pair belonging to the second voice.

Y = the same set as X.

This analysis can be carried out also in multilevel form or for more than two voices. The sets X and Y are then to be modified accordingly.

d. Transition frequency analysis carried out according to the rhythm structure.

In this case several transition frequency matrices are prepared which might in turn belong to any of the kinds of matrices listed above. So the representing graph is transformed into a set of graphs too. The set of axes of the graph $G_i$ represents the transitions in the case if the measures with rhythm structure type i are sorted.

The above list is intended to be only an illustration of the variety of the possible transition analyses which might be planned.

3. Electronic music composition by analysis in rhythm type.

If we have a clear definition of the transition probability matrices, we may use approximative analogue terms in connection with them.

A single-voice music has to be analysed. The tune is formed of musical notes of different pitch and duration following each other and ordered in bars.

Let us suppose for a moment that we regard only the pitch of the notes and we make on the basis of this a one-level or a multi-level Markov-analysis. This has been the general practice in analytic studies till now.

Such a transition analysis can be demonstrated by sorting in different cells, so that the above procedure can be visualised in the following way: we take all the notes which have e.g. "c" as ancestor and we put all these notes in a common cell. The next step is to sort the content of the cells, so that the notes with different pitch are divided in different sub-cells.

If all the notes which have "c" as the foregoing note are put in a common cell, so this method can be regarded only as a very rough approximative one.

In this case it is not taken into account what the duration of the previous note was. Neither is the duration of the actual note sorted.

No distinctions are made either whether the transition from the "c" to another note has been realized crossing the limit of a measure, and this is the case with the distinctive role of the different positions of the measures.

On the other hand, it is evident concerning both the musical effect and the simple observation of the music score that even in the case of the same music piece there are quite different regularities determining the succession of the different notes in a ♩♩♩♩ and in a ♩♩ measure. There are quite different regularities in the transitions, if the ancestor is in a measure and the successor in the next, or if both are in the same measure.

If we put all the notes that are in so different positions but have an ancestor "c", in a common cell, the transition frequencies characteristics for different positions are mixed, and the regularities cannot be detected.

A computer program was prepared to avoid all these insufficiencies.

The main routines in the program are the following:

1./ As input data, the coding of the complete information stored in the music score is required. So we register the pitch, the duration of the note and the limits of the measures.

2./ The computer program is sorting the information got according to the rhythm type /𝅗𝅥, ♩ ♩, ♩. ♪, ♪♩., ♫♩, ♩♫, ♬/. The notes figuring as the first note of a measure or as the last note of a measure are separately sorted, and it is also registered what is their duration and that of their ancestors. The pitch of the notes forms the basis of every type of sorting; so this too is registered.

3./ A set of transition frequency matrices are constructed by the computer program. These are the following:

a/ Such transitions as are connected with a bar transition are registered in four different matrices. The four matrices are distinguished according as the ancestor note or the successor note had a duration of 1/8 or longer. The matrices themselves contain the transitions sorted by the pitch of the ancestor and successor note.

b/ For all types of measure with a different rhythm structure one or more transition frequency matrices are prepared. The basis of the preparation are the note-sets which appear in the analysed music pieces in that type of measure for which the transition matrix is prepared.

In one given matrix sorting is done on the basis of the pitch of the ancestor and successor note. For example, for the measure with the rhythm structure: |  |, two transition matrices are prepared. The first transition matrix will contain information on how often a "c" in the first place of the measure is followed by "c", "d", "e" etc. /In the same way for the "e", "d", etc. figuring as the first note of the measure./ /See Appendix, Fig. 2./.

In the second matrix a two-level transition analysis is made. We register here in what distribution the notes figure in the third /and last/ position of the measure, following on the two element series of notes figuring in the first two places.

Transition matrices for the measures with different rhythm structure from the one discussed above are prepared in a similar way.

If among the measures with a different rhythm structure taken into account there are "x" different types which have two notes, and there are "y" different types which have nore than two notes, then

$$x + 2y + 4$$

transition frequency matrices are prepared.

If there are "n" different pitch levels of notes taken into account, the dimensions of the matrices are the following:

| No. of matrices | Dimension | Position of notes figuring in the matrices |
|---|---|---|
| 4 | n x n | ancestor: end of the measure<br>successor: first place of the next measure |
| x + y | n x n | ancestor: first position<br>successor: second position of the measure |
| y | $n^2$x n | ancestors: first two positions or second-third positions<br>successor: third or fourth position of the measure. |

## Conclusions.

The presented computer composed music has its origin in the analysis of two-voice children choir songs by Z. Kodály and L. Bárdos /see [2] , [3] ./

First the rhythm of the songs has been analysed. The transition analysis was carried out with pairs of rhythm types of consonant measures. This type of analysis is defined in an exact way in 2.3,c.

The matrix which is got as a result of the analysis made it possible that the rhythm scheme of the two-voice song could be composed. The rhythm scheme of the computer-composed tune presented in Figures 3 and 4 has been generated this way.

The transition analysis concerning the pitch of the notes has been carried out according to the method presented in point 3. A special analysis has been done for the upper and for the lower voice. The transition matrices presented in Fig. 1. and Fig. 2. show clearly that in the music small intervals are frequent. The analysis also shows that the music analysed is a well-composed music in the sense that the transitions do not represent a uniform distribution, but they reflect the definite thoughts of the composer. If we compare Fig. 1. and Fig. 2., it is very easy to see what difference is between the transitions in two measures, differing in rhythm type.

Looking at the music score in Fig. 3. or Fig. 4., before judging it one should take into account that the upper and the lower voice are coordinated only as far as the rhythm coordinates. Moreover, the analysis of the notes did not go deeper than the depth of a bar. Beside this only the links between two consecutive bars have been analysed. So, as it has been stated already, these music pieces do not reflect large compositional intentions, but more, they are built of varied forms of the basic ideas in choral music for children by Kodály and Bárdos.

Nevertheless, the defects which are surely shortcomings if we regard the music pieces from an aesthetic point of view, constitute from a scientific point of view a very important basis for a further comparative investigation. So analysing such compositions the role, the importance of the coordination of rhythm and tonality, of the coordination of different voices, of the coordination within one measure, and going beyond one single measure can be seen.

It would be completely false, however, to think that an aesthetically satisfactory formal representation could be attained in the way presented here and in other sources. Such straightforward methods as sorting by different obvious characteristics, the increasing of the depth of Markov analysis, will never yield a complete picture of music.

As these methods are only good for a first orientation, they are incapable of grasping the structural interactions of constitutive elements, to weight these elements against one another, i.e. to find the "grammar" of music.

Such results might be hoped for from a different line of attack, as pointed out by the present author in 6 . This way points to the development of techniques which is able to simulate music concept formation.

x x x

The author is indebted to thanks to K. Csébfalvi, Head of the Computing Centre, Ministry of the Heavy Industry, for making available computer time on the NE 803/B.

Thanks are also due to M.Havass for helping in program debugging and running.

## Appendix

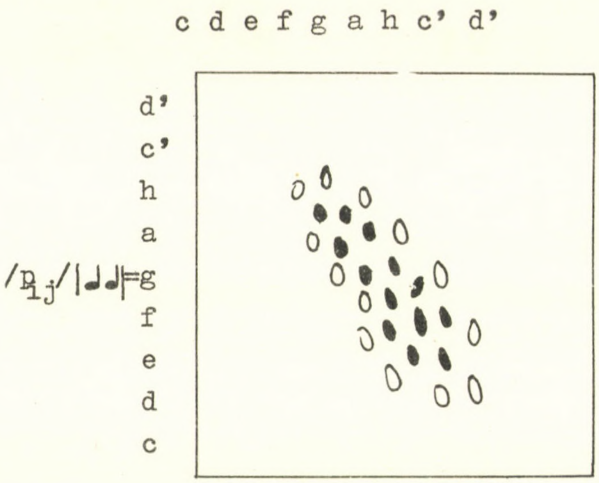



Fig. 1.

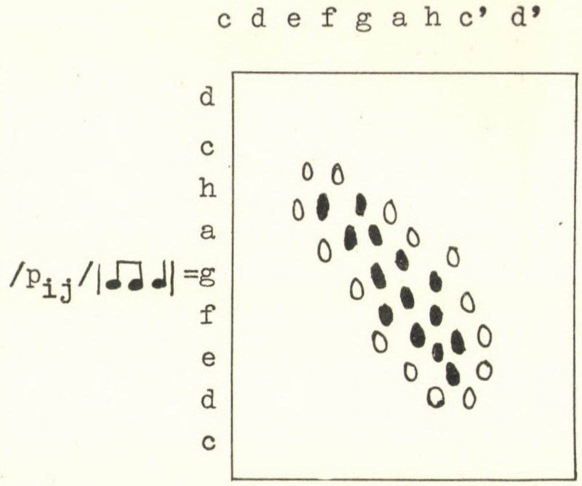



Fig. 2.

Probability matrices of the transitions from the 1st note to the 2nd note in | ♩ ♩ | and | ♫ ♩ | measures, respectively.
Empty cells represent $p_{ij} = 0$ .

Fig. 3

Fig. 4.

## Literature

[1] W. Cooper: The voice in mutation. /A paper presented at the ISME VI. annual Conference, Budapest, 1964/

[2] L. Bárdos: Kicsinyek kórusa. /Choir songs for children. Budapest, 1951. Zeneműkiadó/

[3] Z. Kodály: Kórusok /Choir-songs. Budapest, 1949. Magyar Kórus./

[4] L.A.Hiller, Jr. and L.M.Isaacson: Experimental music. /Mc Graw Hill Co. 1959./

[5] W.Fuchs: Mathematische Analyse von Formalstrukturen von Werken der Musik. /Westdeutscher Verlag, Köln und Opladen./

[6] E.N.Ferentzy: Computer simulation of human behaviour and concept formation in music composition. /Computational Linguistics IV. Budapest 1965. Computing Centre of the Hungarian Academy of Sciences./

[7] K.Csébfalvy - M.Havass: A direct computer processing of folk tunes. /Computational Linguistics IV. Budapest, 1965. Computing Centre of the Hungarian Academy of Sciences./

[8] M.Havass: A simulation of the music composition. Electronically composed folk-music. /Computational Linguistics III. Budapest 1964. Computing Centre of the Hungarian Academy of Sciences./

[9] E.N.Ferentzy and M.Havass: Human movement analysis by computer. Electronic choreography and music composition. /Computational Linguistics III. Budapest, 1964. Computing Centre of the Hungarian Academy of Sciences./

## A direct computer processing of folk tunes.

K. Csébfalvi and M. Havass

In analysing the rhythms and melodies of a gret number of folk tunes we made use of the advantages offered by a high speed, program conducted electronic digital computer. We tried to employ the computer, because our problem consisted in processing information distributed along the melody, which is closely similar to other problems regularly treated with the help of electronic computers.

The computers may be used effectively for the purpose of statistical analysis, from the simplest tasks of classification to complex mathematical problems. Their use requires a very precise and unambiguous statement of all the data and all the criteria of classification, as well as the setting up of a mathematical algorithm for the computation. This demand for a high level of accuracy means of course vigorous requirements in the first stage of work. Every possibility must be taken into account because the computer can perform only exactly defined tasks explicitly stated in the program. However, this increased accuracy means not only additional effort, but at the same time may lead to additional scientific results, valuable even without regard to the application of the computer.

The classification and the analysis of folk tunes presupposes their notation in a unique and unambiguous way, invariant with the personality of the performer. These problems are, of course, far from being new, but the use of the mathematical tool raises them again in a very pronounced way. The resulting additional effort is therefore far from being wasted, because these questions need clarification even without the requirements raised by these methods.

After having found a workable solution to the above questions, we had to store the melodies previously defined unambiguously in the memory of the computer.

The first and obvious possibility was to use the traditional way, and to punch the rhythms and melodies by way of a code system onto tapes or cards. This method, however, is not to be recommended, partly because of the great amount of labour required by this method.

As to the first consideration, information theory states that the errors committed in the course of coding and input processes would result in so-called "noises" which would result in inaccuracies in the course of further analysis. The "noise" factor could only be eliminated in a later stage by the use of a high frequency generator /"audio--control"/. This device reproduces the melodies stored in the memory and the errors can be detected simply by listening. Such experiments have proved successful, but the procedure is rather uneconomical.

The other method is based on a characteristic of the synchron computers, namely that the duration of an impulse may be measured with great accuracy within a given interval.

By employing a bistabile multivibrator system, one channel to each given pitch level, the impulses of different length, originated in whatever channel can be registered in an unambiguous way. /This can be realized by means of the number generator and accumulator of the computer, and by instructions connecting the above units, and counting the number of identical operations performed./ The different multivibrators may be actuated through the resonators - at least in principle. In this case the input of melodies can be carried out through a microphone by means of a well-defined instrument. As, however, such an input requires the solution of a number of technical problems, we actuated the multivibrators by means of an electric connection linked with the keys of the keyboard of a music instrument /a specially developed variant of the piano./ In this way it was possible to play the melodies, to control them by listening and to store them in the memory of the computer.

In our experiment we used a keyboard with three octaves. The sounds and chords were read into the computer in the way described above, and translated into machine language by way of an appropriate program stored in the computer. /The computer actually used is an ELLIOTT 803/B./

The method also makes it possible to switch over from the absolute to the relative periods by way of a suitable program. The absolute periods express the inexactitute of the rhythm and the fluctuation of the tempo with which the melodies are played on the keyboard. The bar lines and measures necessary to the transformation may be put in either by means of special keys of the keyboard, i.e. simultaneously with the playing of the melodies, or by means of a punched tape.

The normalization by bars and the correction within the bars can be assured by the appropriate selection of these bar lines and measures. It is apparent that even in case of the most rigorous keeping of the bar, any rhythm unit will differ both from all other actual rhythm units and from the theoretically computed ones. So, for instance, if the theoretically computed length of a 1/4 sound is denoted by $x_o^4$ /which, of course, depends on the measure type, so, for instance, we have in the case of a three-in-a-measure bar $x_o^4 = 1/3$, and in the case of four-in-a-measure bar $x_o^4 = 1/4$ /then we have for each input quarter

$$\mathcal{E}^4_{1o} = |x_1^4 - x_o^4| \qquad /i = 1,2, \ldots, n/$$

further

$$\mathcal{E}^4_{ik} = |x_i^4 - x_k^4|$$

where

$$\mathcal{E}^4_{ik} \geq 0$$

/i,k = 1,2,..., n/; if n stands for the number of the input quarters/.

/The exponent 4 /generally j/ indicates that these $\mathcal{E}^4_{ik}$ values /in general $\mathcal{E}^j_{ik}$ / are the errors belonging to the quarter sounds /in general to the j-th sounds/.

Now the following question arises: what is the value of the threshold number $\mathcal{E}^4$ for which in the case of $|x_i^4 - x_o^4| \leqq \mathcal{E}^4$ /i = o,1, .. , n/ $x_i$ can be considered as a quarter.

The number may be determined by two different methods:

1./ The $\varepsilon^j$ values may be determined a priori, on the basis of some general experience and knowing the theoretical values of the different rhythm units. That means, we can give an interval $I^j \ni x_o^j$ /not necessarily symmetrical/ in an <u>a priori</u> way, for $x_i$ may be regarded as a $x_o^j$ rhythm unit, if $x_i \in I^j$.

2./ The values of $\varepsilon^j$ may be determined by the computer itself in an a posteriori way, i.e. after a suitable learning period. In this case each playing person must play for a certain period. After playing some standard melodies we obtain the characteristic limits of error. By doing so, we make use of the fact that the actual length of the fact that the actual length of the rhythm unit is a probability variable, the distribution of which is characteristic of each individual, and we may carry out the corresponding statistical analysis.

Our experiments have led to the result that even the simpler method /1./ provided the required level of accuracy.

## L i t e r a t u r e

[1] M. Havass: A simulation of the music composition Electronically composed folk-music. /Computational Linguistics III. Budapest, 1964. Computing Centre of the Hungarian Academy of Sciences./

[2] E.N.Ferentzy and M.Havass: Human movement analysis by computer. Electronic choreography and music composition. /Computational Linguistics III. Budapest, 1964. Computing Centre of the Hungarian Academy of Sciences./

[3] E.N.Ferentzy: Computer simulation of human behaviour and concept formation in music composition. /Computational Linguistics IV. Budapest, 1965. Computing Centre of the Hungarian Academy of Sciences./

# A P P E N D I X

С.К. Шаумян и П.А. Соболева:
Аппликативная порождающая модель и исчисление
трансформаций в русском языке.
/Applicative generative model and the determination of transformations in Russian/
Ц.А.Н. СССР Москва 1963 p. 126.

1.1. Soviet scholars dealt with the problems of transformational grammar at a special conference in 1961. 1/ Since then several papers have been published partly on general questions, partly on questions concerning Russian but all of them looking at the possibilities and deficiencies of generative grammars elaborated by N. Chomsky.

The most outstanding work in this field has been done by Šaumjan. He published the first version of his conception about a new generative model in the journal Voprosy Yazykoznaniya. 2/ Almost simultaneously another paper by Soboleva was published on generative grammars which scrutinized Russian sentences on the basis of Šaumjan's ideas. 3/

Subsequently a systematically elaborated model was put forward in a separate booklet jointly by Šaumjan and Soboleva. It was an extended and modified version of the paper submitted to the 5th Congress of Slavists held in Sofia in 1963. Although this booklet cannot be considered a complete and final representation of the model, so far it is the best elaborated one.

1.2. Before examining the new generative model we want to observe by way of introduction that

1. the model in its present form is not completely elaborated as yet, and this fact is repeatedly stressed by the authors. Nevertheless the model even in its present form deserves attention because some of its ideas may be fruitful or at least suggestive for further investigation.

2. Though the model, as almost every model of language, is a deductive one, it mirrors the properties of an underlying language. Šaumjan's model thus reflects - above all - the properties of Russian. As a consequence, it is not always possible to find examples other than Russian ones. Since Šaumjan's model seems appropriate to describe agglutinative languages we shall refer to some Hungarian examples as well.

3. Our criticism is directed against those points that are not formulated clearly enough or raise unanswered questions. We dc not want to compare this model with others or to check it against the data of a natural language.

4. As our present aim is to follow closely only the main ideas of the authors, we have taken likerties to rearrange some parts of the model for the sake of a higher conceptual accuracy.

2.1. In the Introduction of the booklet the authors subject Chomsky's transformational model to criticism from the following points of view:

a./ The transformational model fails to differentiate between the external, linear /rooted in the speech act/ and internal connections of language.

b./ Though the transformations play a central role in the transformational model, they are mostly given by an arbitrary list and their number is not definable.

The applicative generative model fulfils the requirement a./ and "generates any sentence of the language without any application of transformation". Besides, transformations serve as a tool for finding out some invariant properties of the syntactic structures previously generated by the model and as such their number is in each case quite definable.

2.2.1. The basis of the applicative generative model is the operation called application defined in the following way: application is an operation that states connections between symbols, i.e. is a function with two arguments that assigns to the elements X and Y a third element and this third element is the connection. The application is denoted by: $\overleftarrow{XY}$ or $\overrightarrow{YX}$, where the symbol Y stands for the applicator and X for the applicandum.

This operation is not commutative, i.e. $\overleftarrow{XY} \neq \overrightarrow{XY}$ and is not linear, i.e. two symbols between which there is an applicative connection need not necessarily stand next to each other, so that it may happen that $\overleftarrow{XOY}$.

A sequence of symbols such that an applicative connection holds between its elements is called complex.

2.2.2. We may state a so-called applicative dominancy relation between two symbols X and Y between which an applicative connection is valid. Namely, if X stands for the applicandus and Y for the applicator, i.e. if XY or YX, then we say that X dominates Y. This relation is denoted by XY or YX.

This applicative dominancy relation has the following properties:

a./ it is irreflexive, i.e. none of the elements may dominate itself;

b./ it is asymmetric, i.e. XY and XY may not hold at the same time and

c./ it is intransitive, i.e. from XY and YZ the relation XZ does not follow.

3. Elementary classes, elementary complexes

3.1. We denote by N, V, D, A the word classes which are assumed to constitute the vocabulary Q of the language L. These classes are called elementary classes. /Words do not mean word forms but word stems in this model/.

The applicative connection and the applicative dominancy relation is defined in the case of elementary classes in the following way:

D A N V D

The order in which the symbols occur is immaterial, what is important is that the word classes are represented by these symbols.

3.2. Any segment /consisting of one, two, ..., five elements/ of the above sequence of symbols is called the complex belonging to the language L. These - including the elementary classes - are called elementary complexes.

It is easy to show that the complexes consisting of two or more elements may be generated from the elementary classes by applying the application. If we impose the requirement on the generation that the possible applications defined between the elementary classes can only be applied once, we obtain 14 elementary complexes altogether /including the four elementary classes/.

The generation is uniquely determined by the application graph

$$D \rightarrow A \rightarrow N \leftarrow V \leftarrow D$$

/or by the corresponding matrix/.

The elementary complexes are the subgraphs of this graph

| | | | | |
|---|---|---|---|---|
| D, | DA, | DAN, | DANV, | DANVD, |
| A, | AN, | ANV, | ANVD, | |
| N, | NV, | NVD, | | |
| V, | VD, | | | |

/Henceforth we will denote explicitely the applicative connections and the applicative dominancy relations only when they have some bearing on our considerations/.

The graph of the applicative dominancy relation can be given as follows

$$D \longleftarrow A \longleftarrow N \longleftrightarrow V \longrightarrow D$$

3.2. We define over the set of elementary complexes the following equivalence relations:

1. $NV \longleftrightarrow V$
2. $AN \longleftrightarrow N$
3. $DA \longleftrightarrow A$
4. $VD \longleftrightarrow V$

This means for instance that in any complex NV can be replaced by V and vice versa. The performance of this operation is called identical transformation.

If we use the identical transformations to reduce a given complex, the 3rd and 4th equivalence may be carried out independently of the others but the 2nd only after the 3rd had been carried out. E.g.

$$DANV \text{ ——— } ANV \text{ ——— } NV \text{ ——— } V$$

3.2.3. The reduction, i.e. the equivalence relations make for the definitions of the so-called constitutive dominancy relation:

If the complex XY may be reduced by applying the equivalence relations to X, then X constitutively dominates Y and if to Y, then Y constitutively dominates X.

Between the elementary classes we have the following constitutive dominancy relations:

D A N V D 5

The graph of the constitutive dominancy relation is as follows

$$D \leftarrow A \leftarrow N \leftarrow V \rightarrow D$$

On the basis of the graphs even the dominancy relations may be determined by means of matrices.

3.3. Each elementary complex containing the class V is called elementary sentence. /Notice that the elementary sentence is not the same as the simple sentence/.

From the 14 elementary complexes generated above we get 8 elementary sentences, thus

| | | | | |
|---|---|---|---|---|
| N, | NV, | ANV, | DANV, | DANVD |
| | VD, | NVD, | ANVD | |

3.4. The elementary classes are interpreted in the following way:

| | |
|---|---|
| N | substantive |
| V | verb |
| A | adjective |
| D | adverb |

3.4.1. Leaving out of account the question whether the vocabulary of Hungarian contains or must contain the same elementary classes, we shall now take examples also from Hungarian by way of illustration. Let be "kutya" an element of the elementary class N, "szalad" an element of the elementary class V, "dühös" an element of the elementary class A and "igen", "itt" elements of the elementary class D.

The applicative connection between classes can be interpreted approximately in the following way: there is an applicative connection between two classes if by pronouncing an element of one class /this is the applicator/ it is likely to be followed by an element of the second class. That means, we obtain such pairs as

| | | |
|---|---|---|
| igen - dühös | igen dühös | DA |
| igen - szalad | igen szalad | DV |
| itt - szalad | itt szalad | DV |
| dühös - kutya | dühös kutya | AN |

It is more difficult to determine the role that the elements "kutya" and "szalad" play in the applicative connection. The uncertainty of the applicative connection will be explained in 3.4.3.

For the time being, let us accept the following connection

szalad /a/ kutya    szalad /a/ kutya    VN

It is easy to see that there is no corresponding connection between "igen" and "kutya" or "dühös" and "szalad" respectively.

3.4.2. From our chosen elements we may form the following elementary classes:

| | | |
|---|---|---|
| igen, itt | igen dühös, | igen dühös kutya, |
| dühös | dühös kutya, | dühös kutya szalad, |
| kutya, | kutya szalad, | kutya szalad itt |
| szalad, | szalad itt | |

igen dühös kutya szalad,
dühös kutya szalad itt
igen dühös kutya szalad itt

3.4.3. Let us exemplify the two kinds of dominancy on the longest elementary complex:

igen    dühös    kutya    szalad    itt

We can interpret the dominancy relation between two elements between which there holds an applicative connection as follows: that element dominates the other which is more important, which carries more information.

The dominancy relations are different only in the case of the classes N and V : NV. Through this fact the problems connected with the subject-predicate relationship come to the fore.

The applicative dominancy mirrors the procedure of the factual construction of the sentence, while the constitutive dominancy reflects the equivalence relations between complexes.

On the semantic level this corresponds to the difference between the functions of the substantive and the verb. The function of substantives consists in the denotation of things, and the other parts of speech relate to the designated thing only through their connection with the substantive. /In the applicative dominancy relation the class

N represents the absolutely dominating element/. The function of the verb, on the other hand, consists in the communication of a "message"; the other parts of speech refer to the communication only through the verb. /In the constitutive dominancy relation the class V is absolutely dominating element/.

3.4.4. The interpretation of the equivalence relations, the reduction of given complexes and the choice of the 8 elementary sentences can be done easily, and so they are left to the reader.

4. The R-mapping of the elementary classes and the algorithm of their generation

4.o. To designate the functions the elementary classes have in the elementary complexes, new symbols are introduced which are called relators.

| | |
|---|---|
| $R_1$ | designates the relator of the verb |
| $R_2$ | designates the relator of the substantive |
| $R_3$ | designates the relator of the adjective |
| $R_4$ | designates the relator of the adverb |

4.1. On the basis of the above the mapping can be defined in the following way:

The mapping of the class C through the relator R is defined if there exists a class $L_x$ such that any element of it is in an R relation to one or more elements of X. The class $L_x$ is called the R-image of the class X and denoted RX.

4.2. In general an $R_iX$ construction /wnere $R_i = R_1 v R_2 v R_3 v R_4$/ /read: $R_i$ or $R_1$, or $R_2$, or $R_3$, or $R_4$/ may be generated on the basis of the following recursive formula:

1./ $X = V v N v A v D$

2./ $X = R_iX$

This means that if, for example, $X = V$, then by mapping /through $X = R_iX$/ we obtain in the first step the classes $R_1V$, $R_2V$, $R_3V$, $R_4V$, in the second step, say, from $X = R_2V$ the classes $R_1R_2V$, $R_2R_2V$, $R_3R_2V$, $R_4R_2V$ and so forth. The process of generation may be visualized by the following tables:

First step:

| | V | A | N | D |
|---|---|---|---|---|
| $R_1$ | $R_1V$ | $R_1A$ | $R_1N$ | $R_1D$ |
| $R_2$ | $R_2V$ | $R_2A$ | $R_2N$ | $R_2D$ |
| $R_3$ | $R_3V$ | $R_3A$ | $R_3N$ | $R_3D$ |
| $R_4$ | $R_4V$ | $R_4A$ | $R_4N$ | $R_4D$ |

Second step:

| | $R_iV$ | $R_iA$ | $R_iN$ | $R_iD$ |
|---|---|---|---|---|
| $R_1$ | $R_1R_iV$ | $R_1R_iA$ | $R_1R_iN$ | $R_1R_iD$ |
| $R_2$ | $R_2R_iV$ | $R_2R_iA$ | $R_2R_iN$ | $R_2R_iD$ |
| $R_3$ | $R_3R_iV$ | $R_3R_iA$ | $R_3R_iN$ | $R_3R_iD$ |
| $R_4$ | $R_4R_iV$ | $R_4R_iA$ | $R_4R_iN$ | $R_4R_iD$ |

Here by each mapping four R-classes are generated. /$R_1R_iV$ means, for example, the R-classes $R_1R_1V$, $R_1R_2V$, $R_1R_3V$, $R_1R_4V$ ./

In the course of the second step $4^3 = 64$ R-classes have been produced altogether.
The head of the third step is

$R_iR_iV$ $\quad$ $R_iR_iA$ $\quad$ $R_iR_iN$ $\quad$ $R_iR_iD$

or in shorter form

$R_i^2\,V$ $\quad$ $R_i^2\,A$ $\quad$ $R_i^2\,N$ $\quad$ $R_i^2\,D$

/Here $R_iR_iV$ or $R_i^2V$ stands for the classes $R_1R_1V$,... i.e. 16 R-classes altogether./

The constructions generated in the course of the /m+1/st step are given /in short form/ in the following table:

| | $R_i^mV$ | $R_i^mA$ | $R_i^mN$ | $R_i^mD$ |
|---|---|---|---|---|
| $R_i$ | $R_iR_i^mV$ | $R_iR_i^mA$ | $R_iR_i^mN$ | $R_iR_i^mD$ |

The number of constructions is theoretically infinite.
By realizing a construction $R_1 R_1^m X$ the following rewriting rules may be applied

$$R_1 R_1 \dots R_1 = R_1$$

$$R_2 R_2 \dots R_2 = R_2$$

$$R_3 R_3 \dots R_3 = R_3$$

$$R_4 R_4 \dots R_4 = R_4$$

i.e., for instance,

$$R_2 R_2 R_3 R_3 R_3 R_1 R_1 X = R_2 R_3 R_1 X$$

4.3. The interpretation may, of course, differ according to the different languages. The authors' interpretation refers to Russian, so, for instance, according to their interpretation

$R_2V$

substantive formed of a verb

$С_г$

| | | | |
|---|---|---|---|
| $С_г$ | - | ние | пение |
| $С_г$ | - | тель | мыслитель |
| $С_г$ | - | арь | пахарь |

infinitive

инф

and

$R_1R_1V$

the verbal connection of the substantive formed from a verb

($С_в$ $С_г$)

быть учителем
быть в отчаянии
оказывать влияние
................

the verbal connection of the infinitive

($С_в$ Инф)

должен знать
кончил звонить
..............

5. The generalized forms of the elementary complexes

5.o. The applicative relation defined over the elementary classes is valid for the constructions as well. In general,

$$R_4X \qquad R_3X \qquad R_2X \qquad R_1X \qquad R_4X$$

5.1. By the aid of the constructions the simplest complexes can be written in the following form:

| | | | |
|---|---|---|---|
| $R_4X$, | $R_4XR_3X$, | $R_4XR_3XR_2X$, | $R_4XR_3XR_2XR_1X$ |
| $R_3X$, | $R_3XR_2X$, | $R_3XR_2XR_1X$, | $R_3XR_2XR_1XR_4X$ |
| $R_2X$, | $R_2XR_1X$, | $R_2XR_1XR_4X$, | |
| $R_1X$, | $R_1XR_4X$, | | $R_4XR_3XR_2XR_1XR_4X$ |

/By applying $X=R_iX$ these may be further complicated/.

In this way the more general structure of the complexes may be modelled.

All that has been said about the elementary complexes in paragraphs 3.1. - 3.3. is equally valid for the constructions, dealt with above, i.e. the dominancy relations, equivalence relations etc. as well.

As a matter of fact, the elementary complexes may be considered a special case of the constructional complexes on the basis of the following equivalence relations: $R_1V=V$; $R_2N$ N, $R_3A = A$, $R_4D = D$.

For the sake of illustration let us consider some realizations of a given complex, say, of $R_2XR_1X$.

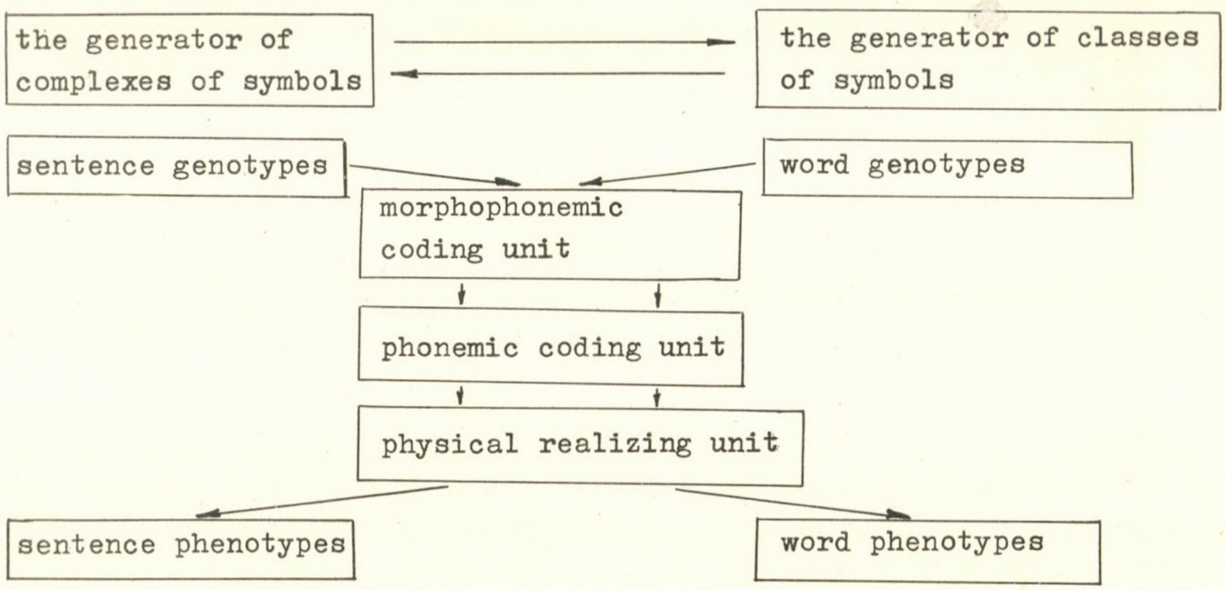
| | | |
|---|---|---|
| $R_2X\ R_1V$ | мумят сады | |
| $R_2X\ R_1A$ | жизнь хороша, | снег был серый |
| $R_2X\ R_1N$ | она студентка, | он был поэтом |
| ...... | | |
| $R_2X\ R_1R_2V$ | он делает заявление, | книга в печати |
| $R_2X\ R_1R_3V$ | книга была открыта | |

Next we shall examine the problems of generating complexes in general.

6. The applicative generative model

According to the authors generation means the production of such grammatical formations that can be built up on the basis of the operations and rules so far established.

The generator may be visualized by the following diagram

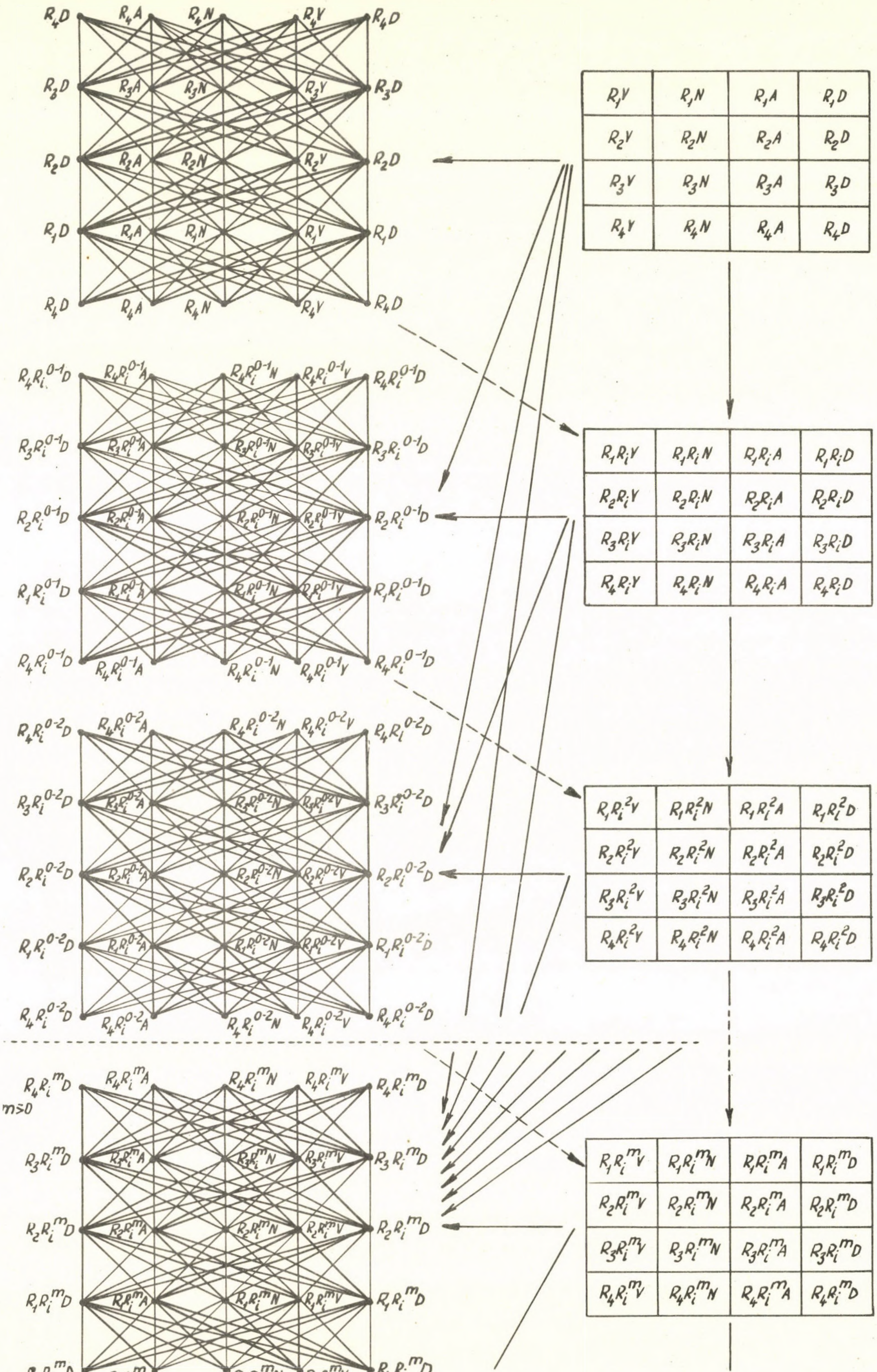



In the abstract genotype unit of the applicative generative model two levels are distinguished: the level of relators and the level of other differentiator. /One part of the grammatical categories belongs to the former, the other part ot the latter level/. In the present form the model must be considered as elaborated only with respect to the genotype part and the level of relators. Other components of the generator are not mentioned explicitly.

7. The model of the paired generator on the level of relators

7.1. The generation of the symbol classes /constructions/ has been shown in 4.2.

7.2. The generation of the symbol complexes is visualized by the graph model connected with the generator of the symbol classes.


$R_4D$ $R_4A$ $R_4N$ $R_4V$ $R_4D$
$R_3D$ $R_3A$ $R_3N$ $R_3V$ $R_3D$
$R_2D$ $R_2A$ $R_2N$ $R_2V$ $R_2D$
$R_1D$ $R_1A$ $R_1N$ $R_1V$ $R_1D$
$R_4D$ $R_4A$ $R_4N$ $R_4V$ $R_4D$
$R_1V$ $R_1N$ $R_1A$ $R_1D$
$R_2V$ $R_2N$ $R_2A$ $R_2D$
$R_3V$ $R_3N$ $R_3A$ $R_3D$
$R_4V$ $R_4N$ $R_4A$ $R_4D$
$R_4R_i^{0-1}D$ $R_4R_i^{0-1}A$ $R_4R_i^{0-1}N$ $R_4R_i^{0-1}V$ $R_4R_i^{0-1}D$
$R_3R_i^{0-1}D$ $R_3R_i^{0-1}A$ $R_3R_i^{0-1}N$ $R_3R_i^{0-1}V$ $R_3R_i^{0-1}D$
$R_2R_i^{0-1}D$ $R_2R_i^{0-1}A$ $R_2R_i^{0-1}N$ $R_2R_i^{0-1}V$ $R_2R_i^{0-1}D$
$R_1R_i^{0-1}D$ $R_1R_i^{0-1}A$ $R_1R_i^{0-1}N$ $R_1R_i^{0-1}V$ $R_1R_i^{0-1}D$
$R_4R_i^{0-1}D$ $R_4R_i^{0-1}A$ $R_4R_i^{0-1}N$ $R_4R_i^{0-1}V$ $R_4R_i^{0-1}D$
$R_1R_iV$ $R_1R_iN$ $R_1R_iA$ $R_1R_iD$
$R_2R_iV$ $R_2R_iN$ $R_2R_iA$ $R_2R_iD$
$R_3R_iV$ $R_3R_iN$ $R_3R_iA$ $R_3R_iD$
$R_4R_iV$ $R_4R_iN$ $R_4R_iA$ $R_4R_iD$
$R_4R_i^{0-2}D$ $R_4R_i^{0-2}A$ $R_4R_i^{0-2}N$ $R_4R_i^{0-2}V$ $R_4R_i^{0-2}D$
$R_3R_i^{0-2}D$ $R_3R_i^{0-2}A$ $R_3R_i^{0-2}N$ $R_3R_i^{0-2}V$ $R_3R_i^{0-2}D$
$R_2R_i^{0-2}D$ $R_2R_i^{0-2}A$ $R_2R_i^{0-2}N$ $R_2R_i^{0-2}V$ $R_2R_i^{0-2}D$
$R_1R_i^{0-2}D$ $R_1R_i^{0-2}A$ $R_1R_i^{0-2}N$ $R_1R_i^{0-2}V$ $R_1R_i^{0-2}D$
$R_4R_i^{0-2}D$ $R_4R_i^{0-2}A$ $R_4R_i^{0-2}N$ $R_4R_i^{0-2}V$ $R_4R_i^{0-2}D$
$R_1R_i^2V$ $R_1R_i^2N$ $R_1R_i^2A$ $R_1R_i^2D$
$R_2R_i^2V$ $R_2R_i^2N$ $R_2R_i^2A$ $R_2R_i^2D$
$R_3R_i^2V$ $R_3R_i^2N$ $R_3R_i^2A$ $R_3R_i^2D$
$R_4R_i^2V$ $R_4R_i^2N$ $R_4R_i^2A$ $R_4R_i^2D$
$R_4R_i^mD$ $R_4R_i^mA$ $R_4R_i^mN$ $R_4R_i^mV$ $R_4R_i^mD$
$m>0$
$R_3R_i^mD$ $R_3R_i^mA$ $R_3R_i^mN$ $R_3R_i^mV$ $R_3R_i^mD$
$R_2R_i^mD$ $R_2R_i^mA$ $R_2R_i^mN$ $R_2R_i^mV$ $R_2R_i^mD$
$R_1R_i^mD$ $R_1R_i^mA$ $R_1R_i^mN$ $R_1R_i^mV$ $R_1R_i^mD$
$R_4R_i^mD$ $R_4R_i^mA$ $R_4R_i^mN$ $R_4R_i^mV$ $R_4R_i^mD$
$R_1R_i^mV$ $R_1R_i^mN$ $R_1R_i^mA$ $R_1R_i^mD$
$R_2R_i^mV$ $R_2R_i^mN$ $R_2R_i^mA$ $R_2R_i^mD$
$R_3R_i^mV$ $R_3R_i^mN$ $R_3R_i^mA$ $R_3R_i^mD$
$R_4R_i^mV$ $R_4R_i^mN$ $R_4R_i^mA$ $R_4R_i^mD$

The nodes of the graph have been connected on the basis of the application, i.e. only those nodes are connected between which an applicative relation holds.

If we depart from any node of the graph in any direction whatsoever and write down the R classes found in the nodes along which we proceed, then we obtain such a complex that may belong to different levels of mapping but are connected with each other by the application. Such a complex may contain more than one element from a given column or row. E.g.

$R_3R_1^mN\ R_2R_1^mN\ R_1R_1^mN$ if $m=0$ Сын инженера - студент

$R_3R_1^mN\ R_2R_1^mV\ R_3R_1^mD$ $m=0$ Вчерашний запуск спутника

7.3. Let us point to some problems in connection with this generating graph.

7.3.1. The fact that we may start it is possible to generate the inverse of each generatable complex /by inverse we mean a complex whose elements are connected by application but the order of their elements is reversed./ However, we do not know anything about the interconnection between complexes and inverse complexes and there is no reference in the authors to the problem as to how they can be transformed into each other.

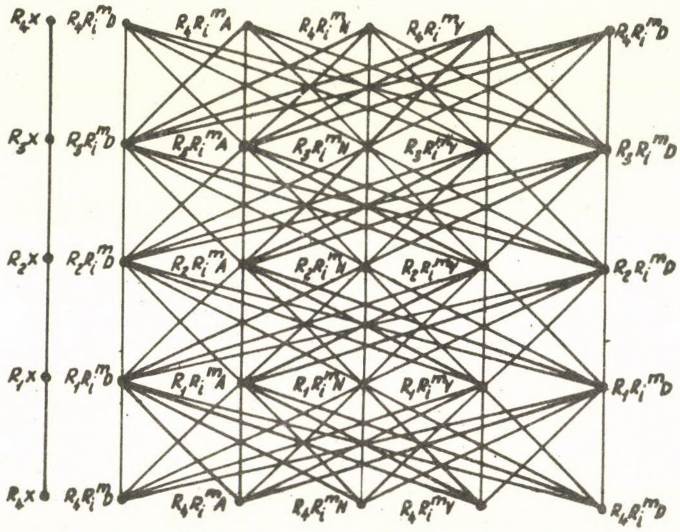
7.3.2. If we consider a given complex and its inverse as identical then it is possible to generate from each complex containing the class $R_1R_1^mD$ four identical complexes because the two outermost columns of the graph are identical. The amplification of the generator by two $R_1R_1^mD$ columns can only be explained by the principle of symmetricity though this is not explicitly mentioned by the authors. This principle plays a role even in the course of the transformation of complexes and so the graphs of these complexes are noticeably the subgraphs of the general graph.

7.3.3. The examples cited by the authors, thus among others $R_3NR_2VR_3D$, point to the fact that a complex may contain more than one R class from one row. Here the following question may be raised.

It is possible from one node to return to the previous one, i.e. is it possible to generate the complex $R_3N\ R_2VR_3N$ ? If so, is this identical to the complex $/R_3N/^2R_2V$ / Cf. 8.2.1./. If the answer is in the affirmative then what are the rules of the transformation?

8. The rules of the production of complexes

8.1. As pointed out in 7.2. it is possible to generate an infinite number of complexes by means of a paired generator. The booklet deal only with the generation of the simplest type of complexes described in 5.1. and with their amplification and transformation. On the basis of the graph model we can obtain the realizations of these constructive complexes and can compute their numbers in the following way. We draw the graph of the constructions next to the /m+1/st graph of the model.



Now we can obtain all possible realizations of the subgraphs of the graph on the left-hand side on the basis of the generator graph on the right-hand side. /Notice that our remarks in 7.3.1. and 7.3.2. are valid in this case too./

So, for instance, the realizations of $R_2XR_1X$ are given

| | | | |
|---|---|---|---|
| $R_2R_i^mD\ R_1R_i^mD$ | $R_2R_i^mA\ R_1R_i^mD$ | $R_2R_i^mN\ R_1R_i^mD$ | $R_2R_i^mV\ R_1R_i^mD$ |
| $R_2R_i^mD\ R_1R_i^mA$ | $R_2R_i^mA\ R_1R_i^mA$ | $R_2R_i^mN\ R_1R_i^mA$ | $R_2R_i^mV\ R_1R_i^mA$ |
| $R_2R_i^mD\ R_1R_i^mN$ | $R_2R_i^mA\ R_1R_i^mN$ | $R_2R_i^mN\ R_1R_i^mN$ | $R_2R_i^mV\ R_1R_i^mN$ |
| $R_2R_i^mD\ R_1R_i^mV$ | $R_2R_i^mA\ R_1R_i^mV$ | $R_2R_i^mN\ R_1R_i^mV$ | $R_2R_i^mV\ R_1R_i^mV$ |

by where the m's in the complexes may range over the interval /0,m/. /See, for example, 5.2./

In the case of concrete languages the upper bound to be imposed on m must be determined. It seems however, to depend on the level under consideration as well /language, complex type, R-class/. Possibly this number is not too high.

8.2. The complexes obtained by generation can be amplified in different ways.

8.2.1. One way of amplification is if we drop the requirement according to which the possible applicative connections could occur only once in any complex. /The applicative dominating, or better, the applicative absolutely dominating element is a single one in each complex./. In this way we obtain the following complexes /on the level of constructions/:

$(R_4X)^n R_3X$ $\qquad$ $(R_4X)^n (R_3X)^n R_2X$ $\qquad$ $(R_4X)^n (R_3X)^n R_2X\,(R_1X)^n$

$(R_3X)^n R_2X$ $\qquad$ $(R_3X)^n R_2X\,(R_1X)^n$ $\qquad$ $(R_3X)^n R_2X\,(R_1X)^n (R_4X)^n$

$R_2X\,(R_1X)^n$ $\qquad$ $R_2X\,(R_1X)^n (R_4X)^n$

$R_1X\,(R_4X)^n$ $\qquad$ $(R_4X)^n (R_3X)^n R_2X\,(R_1X)^n (R_4X)^n$

although it is not explicitly mentioned by the authors it is clear that n 1 /in a complex independently of each other/.

These complexes of constructions represent, syntagmas of different degree of complexity and simple sentences. The complex $(R_4X)^n\ (R_3X)^n\ R_2X\ (R_1X)^n\ (R_4X)^n$ is the complete formula of the extended simple sentence.

The realizations of $(R_3X)^n\ R_2X$ are, for instance,

шерстяной пиджак в клетку $\qquad$ or

старый товарищ брата

The realizations of the complex $R_1X\ (R_4X)^n$

Шел как-то вечером по улице из театра домой $\qquad$ or

выкладывал ему фрукты из шапки в сумку

8.2.2. This sort of amplification, however, does not allow for the expansion of the sentence "Дом стоит в лесу " of type into the sentence "Дом стоит в очень красивом лесу"

However, this may be done by applying equivalence-relations.

The sentence quoted may be expanded by means of equivalence relations in the following way:

$R_2N\ R_1V\ R_4N$ — Дом стоит в лесу

$R_2N\ R_1V\ R_4(NA)$ — Дом стоит в красивом лесу

$R_2N\ R_1V\ R_4(NAD)$ — Дом стоит в очень красивом лесу

Or let us consider the derivation of a more complicated syntagma, namely the syntagma "Далекий от понимания" from the syntagma

which is of type $R_3X\ R_4X$,

We obtain the following derivation

$R_3X\ R_4X$

$R_3A\ R_4(R_2V)$ — далекий от понимания

$R_3A\ R_4(R_2V\ (R_3X))$

$R_3A\ R_4(R_2V\ R_3(R_2N))$ — далекий от понимания целей

$R_3A\ R_4(R_2V\ R_3(R_2N\ R_3X))$

$R_3A\ R_4(R_2V\ R_3(R_2N\ R_3N))$ — далекий от понимания истинных целей

The relators standing before the parantheses refer to the constitutively dominating element of the complex in parantheses.

For the examination of the possibility of the elimination of parantheses the introduction of two notions seems necessary, namely, that of real and of false compound member. By a real compound member is meant a member in the core of which the elimination of the parantheses is not possible because this would lead to the destruction of the existing dominancy relation, in other words it would lead to the establishment of a new dominancy relation.

Such a member is, for instance $R_1R_2N\ R_4(R_1V\ R_4N)$. We can satisfy ourselves of the truth of this statement by comparing the dominancy relations:

$$R_1R_2N\ R_4(R_1V\ R_4N) \qquad\qquad R_1R_2N\ R_4R_1V\ R_4N$$

The complexity of the complex $R_2R_2N\ R_3\ R_1V\ R_4N$ is however not a real one because

$$R_2R_2N\ R_3(R_1V\ R_4N) \qquad\qquad R_2R_2N\ \ R_3R_1V\ R_4N$$

9. The rules of the transformation of complexes

9.1. Definitive transformation or simple transformation.

Transformation is defined in the following way: Transformation means the determination about the invariant connection between two complexes, A and B, where B /the transform/ is derived from A /the operandus/ if the following conditions are fulfilled:

1./ To each $X_i$ class of the complex A it is possible to make a class $X_j$ of the complex B correspond on the same level of mapping and vice versa. /This means that $X_i$ and $X_j$ must be the R-classes of the same elementary class./

2./ If a dominancy relation holds between the classes $X_i$ and $Y_i$ of the complex A, then there must be similarly a dominancy relation between the corresponding classes $X_j$ and $Y_j$ of the complex B but the type of the dominancy relation may be different.

For the sake of illustration let us consider the simplest complexes. Possible interpretations

$R_2N \qquad R_1V \qquad R_4N$ Украина поставляет хлеб

$R_4R_2N \qquad R_1R_1V \qquad R_2R_4N$ Хлеб поставляется Украиной

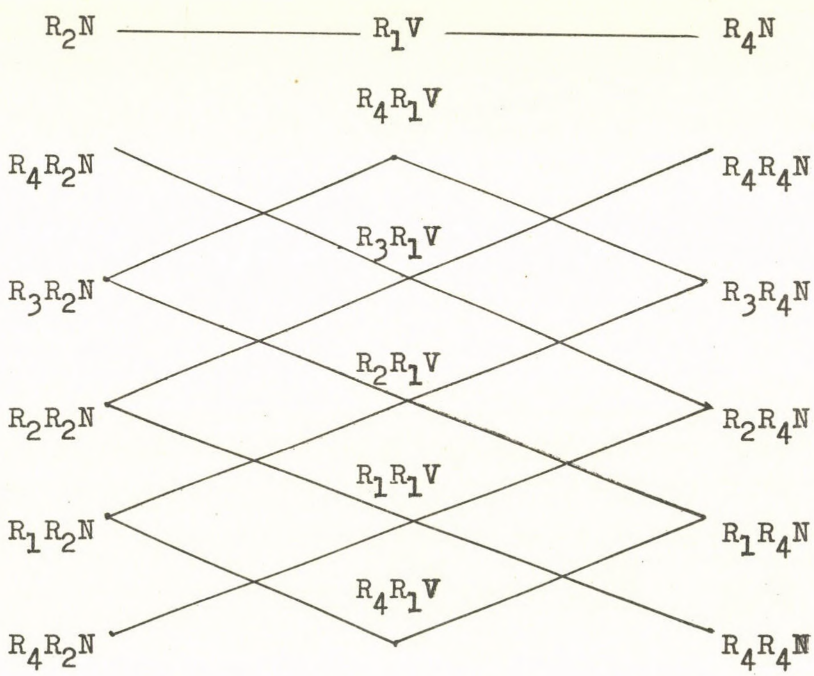
$R_4R_2N \qquad R_3R_1V \qquad R_2R_4N$ Хлеб поставляемый Украиной

This definition of the transformation, the authors think, is only satisfactory on the abstract grammatical level. On a concrete level the conditions must be formulated with respect to the elements $x_i, x_jy_i, y_j$ of the classes $X_i$, $X_j$, $Y_i$, $Y_j$.

9.1.1. The algorithm of the transformation

1. First the $R_i R_i^m$ class of the elements of the operandus must be generated.
2. Let us connect the classes $R_i R_i^m$ belonging to neighbouring elements of the operandus according to the possible applicative relations.
3. Let us take all possible transforms from the graph obtained in this way /an element our of each column, cf. 9.1.5./.

Let us take the transforms of the preceding example in the case of m=0.




It is easy to see that this graph generates 14 transforms. These are the transforms of the first degree of the original complex.

1./ $R_2R_2N$ $R_1R_1V$ $R_2R_4N$

2./ $R_2R_2N$ $R_1R_1V$ $R_4R_4N$

3./ $R_4R_2N$ $R_1R_1V$ $R_2R_4N$

4./ $R_4R_2N$ $R_1R_1V$ $R_4R_4N$

5./ $R_1R_2N$ $R_2R_1V$ $R_1R_4N$

6./ $R_1R_2N$ $R_2R_1V$ $R_3R_4N$

7./ $R_3R_2N$ $R_2R_1V$ $R_1R_4N$

8./ $R_3R_2N$ $R_2R_1V$ $R_3R_4N$

9./ $R_2R_2N$ $R_3R_1V$ $R_2R_4N$

1o./ $R_2R_2N$ $R_3R_1V$ $R_4R_4N$

13./ $R_1R_2N$ $R_4R_1V$ $R_1R_4N$

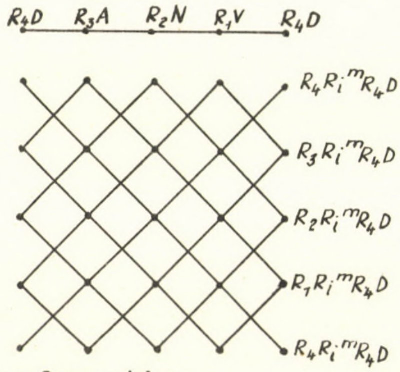
14./ $R_3R_2N$ $R_4R_1V$ $R_3R_4N$

11./ $R_4R_2N$ $R_3R_1V$ $R_2R_4N$

12./ $R_4R_2N$ $R_3R_1V$ $R_4R_4N$

/This is the authors' grouping. Concerning the interpretation of 9.1.5./.
For m=1,2,... we may obtain the transform of the second, third, ... degree.

In general, we obtain the transforms of the /m+1/st degree of the complex $R_4D$ $R_3A$ $R_2N$ $R_1V$ $R_4D$ on the basis of the following graph




9.1.2. Elementary transformation

The determination of the transforms of complexes with several elements may be reduced to the sum of transforms of complexes with two elements.

If we decompose the complex $R_4D$ $R_3A$ $R_2N$ $R_1V$ $R_4N$ into the complexes $R_4D$ $R_3A$ $R_3A$ $R_2N$, $R_2N$ $R_1V$, $R_1V$ $R_4D$ with two elements and add up their transforms according to the rule

$$AB + BC + CD + DE = ABCDE$$

where A,B,C,D stand for the R-mappings of the classes of the operandus, we obtain all the transforms of the original complex.

The transformation of the complexes with two elements

9.1.3. The formation of transforms with compound members

9.1.3.1. According to the equivalence relations we can bring together the complex classes of $R_2NR_1VR_4N$ in all possible ways.

Accordingly we obtain

a./ $R_2N$ $R_1V$ $R_4N$

b./ $R_2N$ $R_1V$ $R_4N$

c./ $R_2N$ $R_1V$ $R_4N$

The complexes obtained in this way are, then, transformed in a way described in 9.1.1. From the operandus $/R_2NR_1V\ /\ R_4N$ we obtain the following transforms:

$$\begin{array}{lll}
R_4R_i^m(R_2N\ R_1V) & R_4R_i^m\ R_4N & R_4R_i^m(R_2N\ R_1V)\ R_3R_i^mR_4N \\
R_3R_i^m(R_2N\ R_1V) & R_3R_i^m\ R_4N & R_3R_i^m(R_2N\ R_1V)\ R_4R_i^mR_4N \\
R_2R_i^m(R_2N\ R_1V) & R_2R_i^m\ R_4N & \ldots\ldots\ldots\ldots\ldots\ldots \\
R_1R_i^m(R_2N\ R_1V) & R_1R_i^m\ R_4N & \\
R_4R_i^m(R_2N\ R_1V) & R_4R_i^m\ R_4N &
\end{array}$$

The transforms of b./ may be generated in a similar way, the complexes under c./ however through a simple $R_1R_i^m$ mapping, i.e.

$$R_4R_i^m(R_2N\ R_1V\ R_4N)$$
$$R_3R_i^m(R_2N\ R_1V\ R_4N)$$
$$\ldots\ldots\ldots\ldots\ldots$$

It is clear by now that this way the number of the transforms of the operandus can be increased ad libitum.

9.1.3.2. Another way to increase the number of transforms presents itself if the classes are brought together only after an elementary transformation have been carried out and the complexes are subsected to further transformations. In our example, according to 9.1.3.1. we will have

a./ $\quad R_2N$ ———— $R_1V$ ———— $R_4N$

The transforms of the complex $R_2NR_1V$ obtained by an elementary transformation are considered as an $/E_j/$ element /the reviewer's notat ion/ and then the elementary transformations of the complexes $/E_j/R_4N$ are formed in a way that the element $/E_j/$ is mapped once.

$$E_j = \left\{ \begin{array}{ll}
R_2N & R_1V \\
R_4R_i^m\ R_2N & R_4R_i^m\ R_1V \\
R_3R_i^m\ R_2N & R_3R_i^m\ R_1V \\
R_2R_i^m\ R_2N & R_2R_i^m\ R_1V \\
R_1R_i^m\ R_2N & R_1R_i^m\ R_1V \\
R_4R_i^m\ R_2N & R_4R_i^m\ R_1V
\end{array} \right\}
\qquad
\begin{array}{ll}
 & R_4N \\
R_4\,(E_j) & R_4R_i^m\ R_4N \\
R_3\,(E_j) & R_3R_i^m\ R_4N \\
R_2\,(E_j) & R_2R_i^m\ R_4N \\
R_1\,(E_j) & R_1R_i^m\ R_4N \\
R_4\,(E_j) & R_4R_i^m\ R_4N
\end{array}$$

This way we obtain the following transforms

$$R_4\ E_j\ R_3R_i^m\ R_4N$$
$$R_3\ E_j\ R_4R_i^m\ R_4N$$
$$R_3\ E_j\ R_2R_i^m\ R_4N$$
$$\ldots\ldots\ldots\ldots$$

Since the number of the elementary transformations amounts to 8, each row will contain 8 transformations. If we expand the first row we obtain

$$R_4\ (R_4R_i^m\ R_2N\ R_3R_i^m\ R_1V)\quad R_3R_i^m\ R_4N$$
$$R_4\ (R_3R_i^m\ R_2N\ R_4R_i^m\ R_1V)\quad R_3R_i^m\ R_4N$$
$$R_4\ R_3R_i^m\ R_2N\ R_2R_i^m\ R_1V\quad R_3R_i^m\ R_4N$$

b./ According to the above first we can form the transforms of the complex $R_1VR_4N$ too, and then the transforms of the complexes $R_2N$ /$E_k$/.

c./ A further possibility presents itself by the mapping through relators, i.e. we may map the transforms of the complex $R_2N\ R_1V\ R_4N$ by means of relators. Then, we obtain

$$R_4\ (R_4R_i^m\ R_2N\ R_3R_i^m\ R_1V\quad R_4R_i^m\ R_4N)$$
$$R_4\ (R_4R_i^m\ R_2N\ R_3R_i^m\ R_1V\quad R_2R_i^m\ R_4N)$$

The question how the compound transforms of complexes with more than three elements can be formed remains entirely unanswered.

9.1.4. If we examine the alterations of the dominancy relations in the course of the transformation we can state that at least one of the dominancy relations is altered in the case of transforms of the first degree, except for the identical transformation.

$$R_4R_2N\quad R_1R_1V\quad R_4R_4N$$

| | | |
|---|---|---|
| K | + | + |
| A | - | + |

K: constitutive

$$R_4R_2N\quad R_1R_1V\quad R_2R_4N$$

| | | |
|---|---|---|
| K | + | + |
| A | - | - |

$$R_2N\ R_1V\ R_4N$$

A: applicative $R_4R_2N \quad R_3R_1V \quad R_2R_4N$

| | | |
|---|---|---|
| K | + | – |
| A | – | – |

$R_2R_2N \quad R_3R_1V \quad R_4R_4N$

| | | |
|---|---|---|
| K | – | + |
| A | + | + |

9.1.5. The authors deal at length with the example given in 9.1.1., namely with $R_2NR_1VR_4N$ and with its simple and compound transforms of the first and second degree. The transformations and the possible interpretations with respect to Russian are summarized in a table taking up 8 pages.

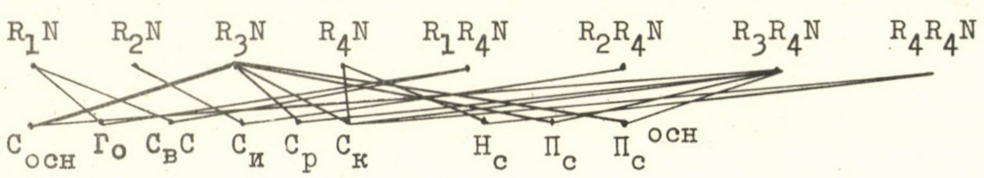
Let us mention here some of these interpretations. The abbreviations and the R classes are interpreted as follows.

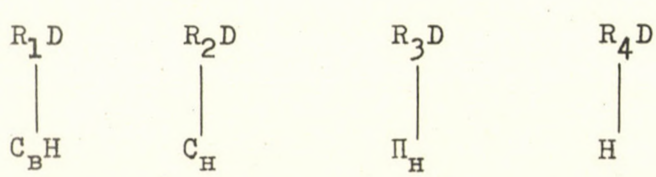
| Термин | Сокращенное обозначение |
|---|---|
| Существительное | С |
| Существительное в именительном падеже | $С_и$ |
| Существительное в родительном падеже без предлога | $С_р$ |
| Существительное в остальных косвенных падежах как с предлогами, так и без предлогов /включая родительный с предлогом/ | $С_к$ |
| Именная основа, входящая в состав сложного слова | $С_{осн}$ |
| Существительное, образованное от прилагательного | $С_п$ |
| Существительное, образованное от глагола | $С_г$ |
| Существительное, образованное от наречия | $С_н$ |
| Субстантированное причастие | $С_{прич}$ |
| Связка /глагольная, неглагольная, нулевая/ | $С_в$ |
| Глагол в личной форме | Г |
| Глагол, образованный от существительного | $Г_с$ |
| Глагол, образованный от прилагательного | $Г_п$ |
| Глагол в форме инфинитива | Инф |
| Глагол в форме деепричастия | Деեп |
| Глагол в форме причастия | Прич |
| Прилагательное | П |
| Отглагольное прилагательное | $П_г$ |

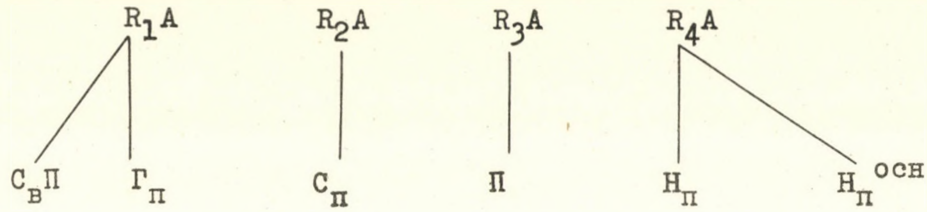
| | |
|---|---|
| Отыменное прилагательное | $П_о$ |
| Прилагательное, образованное от наречия | $П_н$ |
| Основа прилагательного, входящего в состав сложного слова | $П_{осн}$ |
| Наречие | Н |
| Наречие, образованное от прилагательного | $Н_п$ |
| Наречие, образованное от существительного | $Н_с$ |
| Основа наречия, входящая в состав сложного слова | $Н_{осн}$ |

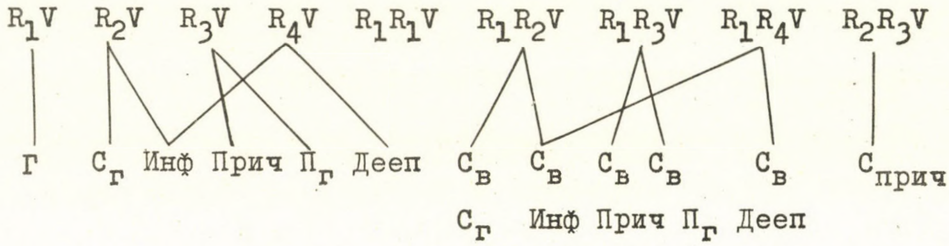
Notice that if there are two indices, one of them is written over the other.

On basis of the above, R-classes:






The operandus $R_2NR_1VR_4N$ rewritten in symbols gives: $С^I_и + Г + \begin{matrix}С^2_к\\ Н^2_с\end{matrix}$

The transforms given in 9.1.1. and their interpretations

6. $$\left\{\begin{matrix}С_вС^I\\ Г^I_с\end{matrix}\right\} + \left\{\begin{matrix}С_г\\ Инф\end{matrix}\right\} + \left\{\begin{matrix}С^2_р\\ П^2_с\\ С^2_к\\ С^2_{осн}\end{matrix}\right\}$$

Петров руководит приедприятием → Руководитель предприятия -- Петров --; Фольга укращает елку → Укращения на елке из фольги; Американцы торговали рабами → Работорговцами были американцы.

II. $$\left\{\begin{matrix}С^I_к\\ Н_с\end{matrix}\right\} + \left\{\begin{matrix}Прич\\ П_г\end{matrix}\right\} + С^2_и$$

Девочка разбила чашку → чашка, разбитая девочкой; Мастер обнаружил изъян → Изъян, обнаруженный мастером; Портниха раскроила платье → платье, раскроенное портнихой; Звезды светят ночью → ночь, светлая от звезд.

To the compound transforms of the first degree of the operandus and of type described in 9.1.3.1.b./ belong the following ones

$R_1R_2N\ R_4/R_1VR_4N/$

$$\left\{\begin{matrix} \text{С}_{\text{в}}\text{С}^{\mathrm{I}} \\ \text{Г}^{\mathrm{I}}_{\text{с}} \end{matrix}\right\} + \left\{\begin{matrix} \text{С}_{\text{г}} \\ \text{Инф} \end{matrix}\right\} + \left\{\begin{matrix} \text{С}^{2}_{\text{к}} \\ \text{Н}^{2}_{\text{с}} \end{matrix}\right\}$$

Путешественник охотится за змеями ⟶ Охотник за змеями --- путешественник /путешествует/;

Солдат бежал из плена --- Беглец из плена был солдатом;

Американцы торговали рабами ⟶ Торговцами рабами были американцы;

Писатели собрались вечером ⟶ Собрание вечером было у писателей.

$R_3R_2N\ R_2/R_1V\ R_4N/$

$$\left\{\begin{matrix} \text{С}^{\mathrm{I}}_{\text{р}} \\ \text{С}^{\mathrm{I}}_{\text{к}} \\ \text{П}^{\mathrm{I}}_{\text{с}} \\ \text{С}^{\mathrm{I}}_{\text{осн}} \end{matrix}\right\} + \left\{\begin{matrix} \text{С}_{\text{г}} \\ \text{Инф} \end{matrix}\right\} + \left\{\begin{matrix} \text{С}^{2}_{\text{к}} \\ \text{Н}^{2}_{\text{с}} \end{matrix}\right\}$$

Человек мечтает о счастье ⟶ человеческая мечта о счастье; Колеса стучат вдали ⟶ стук колес вдали; Голова кружится от успехов ⟶ головокружение от успехов.

Transforms of the second degree are, for instance, the following transforms

$R_2R_2N\ R_1R_2R_1V\ R_2R_4N$

$$\text{С}^{\mathrm{I}}_{\text{и}} + \left\{\begin{matrix} \text{С}_{\text{в}} & \text{С}_{\text{г}} \\ \text{С}_{\text{в}} & \text{Инф} \end{matrix}\right\} + \text{С}^{2}_{\text{и}}$$

Брат спорит с сестрой → Брат и сестра спорщики; Брат ссорится с сестрой → Брат и сестра в ссоре; Коля болтает с Петей → Коля и Петя болтуны.

$R_2R_2N\ R_1R_4R_1V\ R_4R_4N$

$$C_{и}^{I} + \begin{Bmatrix} C_{в}\ \text{Деепр} \\ C_{в}\ \text{Инф} \end{Bmatrix} + \begin{Bmatrix} C_{к}^{2} \\ H_{с}^{2} \end{Bmatrix}$$

Он не выспался после концерта → Он был не выспавшись после концерта; Он стоял в углу → Он оставался стоя в углу.

Transforms of the second degree and with compound members

$$C_{и}^{I} + \begin{Bmatrix} C_{в}C_{г} \\ C_{в}\text{Инф} \end{Bmatrix} + \begin{Bmatrix} C_{р}^{2} \\ C_{к}^{2} \\ П_{с}^{2} \\ C_{осн}^{2} \end{Bmatrix}$$

Он работает на заводе → Он заводской рабочий;
Они убирают сено → Они на сеноуборке; Он воспитывался в Петербурге
Он петербургского воспитания; Он водит автобус → Он водитель автобуса.

$R_1\ /R_3R_2N\ R_2R_1V\ R_3R_4N/$

$$\begin{Bmatrix} C_{р}^{I} \\ C_{к}^{I} \\ П_{с}^{I} \\ C_{осн}^{I} \end{Bmatrix} + \begin{Bmatrix} C_{г} \\ \text{Инф} \end{Bmatrix} + \begin{Bmatrix} C_{р}^{2} \\ C_{к}^{2} \\ П_{с}^{2} \\ C_{осн}^{2} \end{Bmatrix}$$

Ночью светили звезды → Ночной свет звезд.

9.2. Free transformation

The free transformation differs from the transformation in the condition 2./ of 9.1. which is dropped here, in other words the applicative connections may disintegrate.

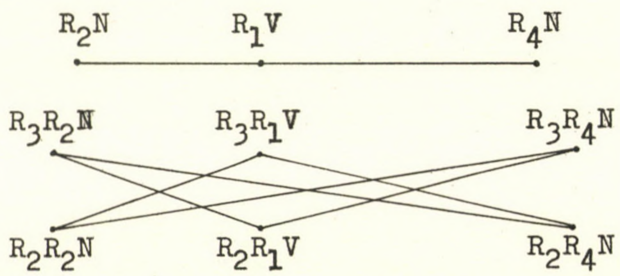
This is a free "conversion" which contains as its special case the transformations as well.

9.2.1. The algorithm of the free transformation

1./ First the classes $R_i R_i^m$ of the elements of the operandus are generated.

2./ The classes $R_i R_i^m$ of any element of the operandus are connected with the classes $R_i R_i^m$ of the other elements according to the application group.

3./ The complexes generatable by the graph are determined by writing out an element of each column. E.g. a part of the graph visualizing the free transformations of the complex $R_2N$ $R_1V$ $R_4N$ has the following form




The complexes obtainable from it

The transforms:

$$\begin{cases} R_3R_2N \; R_2R_1V \; R_3R_4N \\ R_2R_2N \; R_3R_1V \; R_2R_4N \end{cases}$$

Free transformations:

$$\begin{cases} R3R_1V \; R_2R_2N \; R_3R_4N \\ R_3R_1V \; R_2R_4N \; R_3R_2N \\ R_2R_1V \; R_3R_2N \; R_2R_4N \\ R_2R_1V \; R_3R_4N \; R_2R_2N \end{cases}$$

9.2.2. One possible linguistic realization of the previous example

ночной свет звезд
светлые ночь и звезды

звезды светят ночью

светлая ночная звезда
светлая звездная ночь
звездные свет и ночь
ночные свет и звезды

However, only the possibilities of such free transformations are stressed by the authors without any detailed discussion of this topic. The main point with respect to free transformations is the invariancy problem, i.e. what is left invariant in the course of the free transformations.

It is clear that the graph of the free transformation /and consequently that of the transformation/ is a subgraph of the generator.

lo. At the end, the relation of this model to Chomsky's generative grammars is raised once more. Let us quote some of the statements.

On the level of the immediate constituents two levels are involved, a fact that makes the application difficult. Sentences of the type "I have spoken" contain discontinuous morphemes as have...en the realization of which necessitates a separate operation. These two levels are separated by the applicative generative model and labeled as genotype and phenotype and so the model is simple and neatly arranged, and no transformations are needed.

In this model only one operation exists: the application. The transformations do not play any role in the course of the generation, they only serve as a tool for the determination of the interconnections between syntagmas and sentences. They make for the grouping of the transforms into so-called invariancy groups and the establishment of the hierarchy between the transforms belonging to different levels /transforms of the first, second, third, ... degree/.

All these properties as well as the possible transforms with their realizations in given languages can serve as a firm basis for typological research.

## N o t e s

1./ Concerning previous studies and theses see Ferenc Papp, Mathematical Linguistics and Machine Translation in the Soviet Union, Mouton and Co., The Hague /forthcoming/ 2.

2./ С.К. Шаумян, Порождающая лингвистическая модель на базе принципа двухступенчатости В.Я. XII /1963/, 2: 57-71

3./ П.А. Соболева, Опыт исчисления трансформаций на основе теории С.К. Шаумяна о порождении классов слов в процессе порождения грамматики. Проблемы Структурной Лингвистики, М. 1963. pp. 233-265.

4./ The irreflexivity means that no symbol may hold an applicative relation to itself.

5./ It should be stressed that the designation "constitutive dominancy relation" is introduced by the present reviewer. The authors do not give any designation at the beginning where this notion is introduced. Later the same notation is used as in the case of the applicative relation. As this may result in confusion we aimed at the elimination of this ambiguity.

6./ The authors speak in the definition only of sentence and the notion of the elementary sentence is only later introduced.

7./ It should be made clear once again that the elementary classes contain word stems and not word forms.

/This review is the English version of my review written for Studies in General Linguistics IV. /ed. Zs. Telegdi/. There was no possibility here to compare Saumjan's model with models well-known in algebraic linguistics but it is intended to take up this matter in a subsequent paper. Nevertheless it is hoped that this review, even in its present form, will be helpful for those who wish to get acquainted with Saumjan's ideas./

S.J. Petőfi

Papp Ferenc: Matematikai nyelvészet és gépi fordítás a Szovjetunióban. Országos Müszaki Könyvtár és Dokumentációs Központ. Budapest 1964., pp. 222. /Mathematische Linguistik und Maschinenübersetzung in der Sowjetunion/

Dieses Buch mag von grossem Interesse sein für Forscher, die sich mit den neuesten Methoden der Sprachwissenschaft befassen wollen. Obgleich wir uns in den letzten Jahren über die bedeutsamen Erfolge der sowjetischen Sprachwissenschaft aus einigen Studien und Besprechungen wenigstens in grossen Linien Kenntnisse verschaffen konnten, eine ausführliche Darstellung der verschiedenen Richtungen der mathematischen Linguistik in der Sowjetunion liess auf sich warten.
Das Buch von Papp füllt nicht nur diese Lücke aus, sondern ist ein unersetzbares Handbuch für alle, die sich auf diesem Gebiet orientieren wollen.

Im ersten Kapitel wird die Entwicklung der verschiedenen linguistischen Schulen in der Sowjetunion vom Ende des 19. Jhs. bis zur Gegenwart verfolgt. Darauf folgt dann - im zweiten und dritten Kapitel - die Darstellung der Problematik der wohlbekannten Diskussion über den Strukturalismus. Parallel zum Verlauf dieser Diskussion entstanden die Zentren der mathematischen Linguistik in Moskau, Leningrad, Kiew, so wie die Zeitschriften und Periodiken, in denen die ersten diesbezüglichen Publikationen erschienen. Es werden kurz die Themata der internationalen und wichtigeren lokalen Konferenzen angegeben und auf die Stellungnahme gegenüber dem Strukturalismus higewiesen.

Nach der Beschreibung der Entwicklungsgeschichte der mathematischen Linguistik in der Sowjetunion folgt im vierten Kapitel die thematische Darstellung der mathematisch-strukturalistischen Theorien und Probleme. Zu Beginn werden die Arbeiten aufgezählt, die die Probleme der Modellierung, der Anwendung der mathematischen Linguistik in Angriff nehmen. Besonders hervorgehoben werden die Werke von Schaumjan, in welchen man "die Geburt, bzw. die schöpferische Weiterführung einer, die ganze wissenschaftliche Erkennung umfassende Theorie beobachten kann". Es werden die Handbücher der mathematischen Linguistik von Achmanowa, Meltschuk und Padutschewa erwähnt, dann werden wieder die Werke behandelt, die sich mit der Anwendung der

Statistik der Wahrscheinlichkeitslehre und Informationstheorie in der sprachwissenschaftlichen Forschung beschäftigen.

Die Grundprinzipien und Perspektiven der Anwendung von statistischen Methoden fasste Frumkina folgendermassen zusammen: "Die Anwendung von statistischen Methoden führt uns zum Erkennen solcher Gesetzmässigkeiten, die mit anderen Methoden manchmal sehr schwer, andersmal gar nicht zu beschreiben sind."

Im Zusammenhang der Wahrscheinlichkeitslehre wird Toporow zitiert: "Die Einführung der Wahrscheinlichkeit in die Sprachanalyse soll nicht nur als eine ergänzende Methode gelten, sondern auch als Notwendigkeit ... denn alle Sprachmodelle werden die Struktur der Sprache nur annähernd widerspiegeln. Alles was vom Modell nicht oder eben nur im grossen und ganzen und undifferenziert widerspieglet wird, gehört zum Bereich der Wahrscheinlichkeitsverhältnisse."

Die nacheinander aufgezählten Werke liefern einen Beweis zur obigen Feststellung. Diese Werke befassen sich mit Phonemenstatistik, mit statistischer Forschung der Syllabenstruktur, mit dem Vorkommen verschiedener Phonemengruppen, mit der Konstruktion von Frequenzwörterbüchern, mit der Bestimmung der Frequenz grammatischer Formen und Syntagmentypen, mit der Bestimmung der Sprachverwandschaft, mit der statistischen Annäherung der Gesetzmässigkeiten der Sprachentwicklung, mit der Feststellung von Gedichtformeln und mit deren Klassifikation.

In folgenden findet der Leser Fragen über die linguistische Verwendung der Informationstheorie, über die Bestimmung der Redundanz einzelner Paradigmen und Phonemensystems der russischen Sprache. Dieses Thema wird mit einem kurzen Bericht über das im Jahre 1962 abgehaltene semiotische Symposium abgeschlossen. Die Vortragenden dieses Symposiums befassten sich mit den Problemen verschiedener Zeichensysteme, die in der menschlichen Gesellschaft angewendet werden. In einer speziellen Sektion wurden die allgemeinen Fragen, die schriftlichen Zeichensysteme, die künstlichen Sprachen, die aussersprachlichen Kommunkationssysteme, die modellierenden semiotischen Systeme, die Fragen wie Kunstwerke, literarische Schöpfungen mit mathematischen Methoden untersucht werden könnten, besprochen.

Das fünfte Kapitel befasst sich mit einer anderen Richtung der mathematischen Linguistik, nämlich mit den mathematisch-logischen und mengentheoretischen Modellen der Sprache. Im Zusammenhang mit den allgemeinen Problemen wird Rewsin zitiert, der an einer Konferenz, wo er über die Erfolge der statistischen Linguistik sprach, betonte, dass es nur dann möglich sei wesentliche Erfolge zu zeitigen, wenn andersartige Methoden zur Bestimmung der Struktur der Sprache entwickelt werden, da man statistisch nur das bearbeiten kann, was schon "strukturell" beschrieben worden ist.

Unter den mengentheoretischen Modellen wird vor allem das Modell von Kulagina hervorgehoben, das auch von anderen weiterentwickelt und ergänzt wurde /vgl. Rewsins Buch: Die Modelle der Sprache/.

Mit der Transformationsanalyse beschäftigte sich eigens eine Konferenz. Die behandelten Fragen werden auch ausführlich in diesem Kapitel besprochen. Dem Leser wird in die Erfolge der Moskauer und Leningrader phonologischen Schulen, in Schaumjans Theorie, in die Arbeiten auf dem Gebiete der morphologischen und syntaktisch-strukturellen Forschungen sowie in die verschiedenen Fragen der Semantik Einblick gewährt.

Im 6. Kapitel werden Fragen der Maschinenübersetzung und die damit im Zusammenhang stehenden Arbeiten dargestellt. Eine kurze Beschreibung der französisch-russischen, englisch-russischen und ungarisch-russischen Übersetzungsalgorithmen findet man hier ebenso wie Fragen der Ausarbeitung einer Zwischensprache /Interlingua/.

Die letzten zwei Abschnitte beschäftigen sich mit den Verhältnissen der mathematisch-linguistischen Forschung und des Sprachunterrichts und mit den Problemen der Ausbildung von mathematischen Linguisten in der Sowjetunion.

Die wertvolle, aufschlussreiche und nützliche Zusammenfassung wird von einer Bibliographie, die 515 Hinweise enthält, abgeschlossen.

S.J. Petőfi

Roman Jakobson, Selected Writings. T. I -- Phnological Studies. Mouton and Co. 'S-Gravenhague 1962. —— X + 678 pp.

1916-1961. Вот даты написания первой и последней статьи, вышедшей в рецензируемом сборнике избранных работ Р.О. Якобсона. Больше половины века -- и к тому же какого века для лингвистики. Ведь именно за эти десятилетия языкознание пережило свою вторую революцию, если за первую считать появление сравнительно-исторического метода, т.е. становление лингвистики наукой в истинном смысле этого слова. 1916 -- это на год позже появления первого издания работы Соссюра (с которой сам Р.О. Якобсон познакомился уже в 1917-ом году, благодаря С. Карцевскому, см. стр. 631). А последнюю дату можно сравнить с датой выхода в свет пёрвого английского издания Принципов языка (Fundamentals of Language: 1956) Якобсона и Халле, оказывающих огромное влияние особенно в области фонологии полным изложением дифференциальных признаков на основе дихотомической классификации. Уже и сами "внешние" данные сборника внушительные: в опубликованных в нем 37 работах привлекаются материалы более чем из двух сот языков, ссылается на почти семьсот авторов. Некоторые работы автора выходят здесь впервые (так, в этой связи прежде всего надо упомянуть две лекции Р.О. Якобсона, прочитанные им в Копенгагене в мае 1939 г. о структуре фонемы -- Zur Struktur des Phonems -- 280-310). Другие стали уже давно библиографической редкостью -- так, его работы, вышедшие в Трудах Пражского кружка. Наконец, некоторые работы автора, включенные в сборник, были написаны в годы, когда связи нашей лингвистики с наукой запада свелись почти на нет, поэтому мы с ними встречаемся впервые. Правда, есть еще одна группа работ, написанных в последние годы и известных всем, кто поистине интересуется хотя бы судьбами современного языкознания -- но их включение в сборник обосновано именно их выдающейся, основополагающей принципиальной важностью (сюда относится прежде всего отрывок из упомянутых Принципов языка -- стр. 464-504). В результате подбора материала читатель получает полную картину о развитии взглядов Р.О. Якобсона, он имеет в своих руках важнейшие новые исследования автора.

При таком изобилии материала рецензент должен ограничиваться, естественно, только некоторыми замечаниями.

Статьи, помещенные в сборнике, можно было бы разделить на следующие тематические группы: I. Исследование в области общей фонологии, раскрытие понятия фонемы; 2. Исследование вопроса "о языковых союзах"; 3. Исследования из области исторической фонолигии; 4. Исследования из области детской речи и афазии; 5. Исследования из области фонологических структур различных конкретных языков (как-то: русского, словацкого, греческого, идиш, арабского и т.д.); 6. Статья о некоторых вопросах русской графики (Избыточные буквы в русском письме -- стр. 556-567, и этот материал публикуется здесь впервые, одновременно с его появлением в одном сборнике в Югославии). Нельзя забывать, наконец, о трех работах автора, написанных им в молодые годы (о фонетике одного северно-великорусского говора, о некоторых вопросах чешской фонологической системы) и публикуемых здесь в качестве приложения. Том заканчивается замечательным "Ретроспектом" Р.О. Якобсона, подытоживающим "инварианты" в его многолетних поисках в области фонологии (стр. 629-658). Все эти "разделы", произвольно установленные нами, естественно, теснейшим образом связаны между собой и тематически, и, особенно, единой системой, к тому же постоянно усовершенствующейся, взглядов автора на язык.

I. С общетеоретической точки зрения особенно интересен вопрос о том, в чем расходится Р. О. Якобсон с Соссюром, иначе говоря -- чем, в принципиально-теоретическом смысле, он развил дальше нашу науку после женевского мастера. -- Одно из основных открытий, приписываемых обычно Соссюру (мы говорим -- приписываемых, ибо Р.О. Якобсон упоминает в этой связи и фамилию В. Хенри) -- противопоставление языка и речи. Как раз в впервые опубликованных копенгагенских лекциях своих автор указывает на опасность механического понимания этой оппозиции. Здесь скрыты, по его мнению, по крайней мере три противопоставления: между речевой нормой и речевым выражением (sprachliche Norm -- sprachliche Ausserung -- пожалуй, можно было бы по-русски сказать и языковую норму и языковое высказывание), между "центробежным" и "центростремительным" в языке (284-5). -- Со многих сторон и во многих работах освещается Р.О. Якобсоном его отношение к соссюровской оппозиции синхронии-диахронии. Так, уже в 1928-ом году он указывает на то, что языковые изменения не могут быть поняты в отрыве от системы, но и система не может рассматриваться в зависимости от изменений(19). (Автор не избегает здесь еще доли психологизма, присущего и Соссуру: по его дефиниции синхронная система -- это не что иное, как система, существующая в какую-нибудь данную минуту в языковом сознании коллектива говорящих -- см. там же.) Эта же мысль о единстве изучения языковых изменений и синхронной системы языка выступает в Принципах языка (см., напр., 502). В своем докладе на конгрессе лингвистов в Осло Р.О. Якобсон также подчеркивает разницу между "статикой" и "синхронией": "Первоначально все изменения относятся к синхронии: как старый, так и новый вариант встречаются одновременно в одном и том же языковом коллективе, как более архаичный или более современный соответственно..." (528 -- на англ. яз. см. об этом также и в его коппенгагенских лекциях -- 306 и т.д.) Совершенно естественно на этом основании, что Р.О. Якобсон отклоняет известное сравнение Соссюра языка с шахматами, доводит его до конца, указывая на то, что не только правила игры, но и правила замены одной фигуры другой правят структурой языка (531; ср. также стр. 294: наверно, что есть только оппозиция членов.)-- Р.О. Якобсон выступает против "антителеологической" концепции Соссюра так же (327), как и против того, чтобы не учитывать реально существующие иерархические расхождения, напр., в такой системе, как система падежей (525).

Естественно, что система взглядов Р.О. Якобсона характеризуется не только отрицательной чертой -- чего он не признает в Соссюре.

Чтобы начать в этой связи с наиболее общего вопроса об онтологической сущности языка -- Р.О. Якобсон в такой форме отклоняет его, как не лингвистический, а чисто философский (ср. 282). В то же время и он указывает на то, что язык как система -- это объективно необходимая (стало быть, не чисто эпистемологическая) гипотеза, без которой невозможно понимать ежедневно наблюдаемые массовые факты взаимопонимания между людьми с помощью языка. Известно, что в наиболее полном виде эта мысль развита дальше в работах С.К. Шаумяна, сравнивающего язык как абстрактную систему с черным ящикомтеории кибернетики. С.К. Шаумян тем самым и предлагает эксплицитное решение вопроса об онтологической сущности языка, познаваемого лишь путем изучения протокола, составленного на основе анализа входов и соответствующих выходов в черный ящик и из него. [I] Р.О. Якобсон в то же время отклоняет взгляд, как будто понятия лингвистики были введены лишь "для удобства" научного упорядочения материала. Он ссылается на Поста, утверждающего, что понятие оппозиции соответствует аналогичной "мысли", лежащей в самих вещах (423), выступает против "фикционалистического" взгляда Туаделля (472), отвечая на вопрос, поставленный в одной из рецензий Чао он указывает на то, что и дихотомическая классификация фонем присуща самой структуре языка. (Говоря несколько более обычной для нас терминологией -- дихотомическая классификация дифференциальных признаков фонем, предложенная Р.О. Якобсоном, является, по мнению самого автора, не только средством научного познания или упорядочения материала, но и объективным отражением этого куска действительности -- ср. 499-500, с подробным обоснованием этого тезиса.) -- Особенно в своих копенгагенских лекциях Р.О. Якобсон анализирует семиотическую природу языка. В этой связи он указывает на качественную разницу между фонемами и остальными единицами языка (морфемами, словами, предложениями). Фонема -- это элемент знака, сама еще не знак, а все остальные элементы языка, построенные в конечном итоге из фонем -- знаки или комбинации знаков. Тут-то он и подходит к установлению одной из основных черт естественных языков: "Одна из наиболее своеобразных и существенных диалектических антиномий языка заключается в том, что эта наиболее содержательная знаковая система в то же время единственная из этих систем, построенная из пустых, отрицательных основных единиц" (304). [2] Он и разделяет "знаковые стоимости", встречающиеся в языке, с точки зрения процесса обозначения, на классы. Чтобы начать с последнего класса: буквы, явления третьего класса, суть знаки фонем, знаков второго класса, а фонемы -- есть знаки морфем и одновременно: знаки в морфемах , этих знаков содержания первого класса 295 и сл. [3]

Весьма интересно также развитие взглядов Р.О. Якобсона на фонемы. Он восстанавливает истинное значение соссюровских слов о фонемах, как о мельчайших дифференциальных элементах языка -- оказывается, уже в самих лекциях женевского мастера было указание на "дифференцирующие черты" в фонемах. Особенно начиная со своих работ, написанных им в I938-4I гг. (как он на это сам указывает -- см. 435), вырисовывается система дифференциальных элементов фонем. Так, говоря о фонологической классификации согласных, Р.О. Якобсон пишет уже в I938 г.: "Мы разграничиваем фонемы какого-нибудь языка, выделяя мельчайшие звуковые отрезки, способные к смыслоразличению. Мы идентифицируем фонемы какого-нибудь языка, разлагая их на составляющие их фонологические характеристики [caracteres phonologiques constitutifs] (272). В исследовании детской речи и афазии (I939-4I гг.) также вводятся в качестве основных понятий понятие фонемы и "дифференцирующих характеристик" (disinktive Qualität -- см., напр., 388). К сожалению, нет здесь места проследить дальнейший путь развития этой теории у Р.О. Якобсона, которая засвидетельствована рядом интересных работ, публикованных в настоящем сборнике. Логическое упорядочение дифференциальных элементов, а также важнейшие теоретико-информационные выводы из этой теории даны в статье, совместно написанной Р.О. Якобсоном, Э.К. Черри и М. Халле (1953), знаменитые двенадцать пар дифференциальных элементов, больше которых до сих пор не удавалось обнаружить ни в одном из языков мира, наиболее полно изложены в главе Принципов языка (см. здесь стр. 484 сл.).

К 1939-41 гг. относятся первые работы Р.О. Якобсона о детской речи и об афазии. В них также ярко обнаруживаются общефонологические и общелингвистические взгляды автора. Так, неоднократно в этих работах подчеркивается необходимость в системном подходе к явлениям языка. Указывается на развитие от более простого к более сложному в области языка (см., напр., стр. 322). Проводится интереснейшая параллель между развитием речи ребенка и предполагаемым путем развития языка, между развитием речи ребенка и структурной обусловленности отдельных элементов фонологической системы языка. Укажем в этой связи лишь на одну, интереснейшую мысль автора. Тогда как первый гласный и первый согласный у ребенка представляют собой оппозицию в последовательности (полная открытость -- полная закрытость: первыми согласными появляются как раз лабиальные взрывные), появление второго согласного, создавая среди согласных оппозицию ротовые-носовые, представляет собой также первую оппозицию на оси одновременности (носовые -- это как бы одновременное единство известных до сих пор ребенку двух фонем, ведь носовые и полностью открытые и полностью закрытые одновременно).

2. Ряд своих работ 30-х годов посвящал Р.О. Якобсон проблематике языковых союзов; к этим вопросам он как бы возвращается в своем докладе на Конгрессе лингвистов в Осло (1957). Он критикует сравнительно-историческое языкознание прошлого века за его почти исключительное увлечение изучением генетических связей. В то же время много полезного можно извлекать из изучения территориально сопредельных языковых систем, или, наконец, вообще из типологического сопоставления языков. Так, оказывается, в центральной части евразийского материка вырисовываются довольно четко границы одного огромного языкового союза, даже если принять во внимание только одни фонологические черты. "Евразийские" языки объединяются двумя наиболее общими

чертами: отсутствием политонии и наличием различения согласных по "мягкости-твердости". Как раз в этих работах Р.О. Якобсон, между прочим, неоднократно ссылается на венгерский язык, находящийся на самой периферии этого союза, уже вне его пределов и показывающий в свою очередь много общего в своем фонологическом облике в маленьком (по сравнению с евразийским) языковом под-союзе окружающих его индоевропейских языков.

В наших странах, пожалуй, наиболее известны работы Р.О. Якобсона в области общей и с т о р и ч е с к о й ф о н о л о г и и и в частности -- его, ставшееся уже классическим, исследование о развитии фонологической системы русского языка (1930-31 гг. и 1927-28 гг. соответств.). Они читались и читаются, комментировались и комментируются, что освобождает рецензента настоящего сборника от обязанности касаться, хоть и бегло, и их.

Предложения, внесенные Р.О. Якобсоном в у п о р я д о ч е н и е р у с с к о й г р а ф и к и, сводятся к следующему: после ц писать последовательно и (цигане); после шипящих и ц -- о; отменить ъ; отменить ь после шипящих, но пользоваться им и на месте буквы й, при полной отмене последнего /змеь, строь/.

Не хотелось бы, чтобы эти предложения Р.О Якобсона оставились незамеченными как раз в наши дни, когда, как известно, проводится широкая дискуссия по этим вопросам.[4] Это не означает, что мы во всем согласны с автором. В частности -- знак ь для обозначения звука j в позиции после гласного нам кажется целесообразным потому, что этим нарушилась бы функция ь как знака, всегда связанного с предшествующим знаком (т.е. изображаемого вместе с ним как одна буква, "без отрыва карандаша от бумаги"[5]).

3. За статьями, целыми монографиями, отрывками из объемистых работ, помещенных в хронологическом порядке, вырисовывается не только фигура лингвиста, но и ф и г у р а ч е л о в е к а. Первое, что лично нас, как иностранного читателя поразило, что этот человек -- русский до мозга костей, несмотря на всю "космополитичность" его судьбы. Русский язык -- это для него не один из двухсот с чем-то приведенных в качестве иллюстартивного материала языков, он всегда прежде всего и охотнее всего привлекается автором. Работы в сборнике помещены на русском, английском, французском, немецком языках, (статьи, вышедшие первоначально не на одном из вышеупомянутых языков были переведены на английский язык) -- но даже в нерусских по языку исследованиях, лекциях,

хотя бы в виде литературной ссылки, пробивается "Русь". Русский он по своему воспитанию -- одинаково черпал как из достижений фортунатовской, так и казанской школы, ссылается с большим согласием на работы представителей этих направлений и на других русских и советских ученых. (Отсутствует, пожалуй, один только Л.С. Выготский в связи с детской речью, подходивший однако, как известно, с совершенно иной точки зрения к этой проблематике, чем Р.О. Якобсон.) Прямо поразительно, как это удавалось Р.О. Якобсону в век, переполненный такими "катаклизмами" (ср. стр. 632.) ни на минуту не терять связь с русскими учеными, постоянно быть в курсе их творческих интересов, достижений. (Только в одном месте кажется нам, что этот патриотизм отражался и на взглядах автора -- писавшего в 1930 г. о кириллице, как о самом идеальном "евразийском" письме. Ведь слоговой принцип можно было бы, пожалуй, осуществить и на латинской основе, без опасности, что из латиницы остались бы тогда "лишь рожки да ножки" 194.)

4. Надо сделать еще один вывод, вполне практического порядка. Университетские курсы современного русского языка (по крайней мере -- изданные в печати и поэтому известные нам) как у нас, так и в Советском Союзе сильно отстают от современного состояния науки в разделе "фонетика -- фонология" (не будем теперь говорить, естественно, о других разделах этого курса). Р.О. Якобсон в своей рецензии на фонетику современного русского языка Р.И. Аванесова (533-7) совершенно правильно указывает на то, что эта книга, при всех ее достоинствах, страдает тем, что опирается на моторно-артикуляционное определение звуков. Дифференциальные признаки фонем в теории Р.О. Якобсона связаны с акустической стороной, так как последняя с помощью современных технических средств совершенно экзактно измеряется, даже наглядно (в буквальном смысле этого слова) вырисовывается. Хотя, конечно, никакая фонетическая "контрабанда" в фонологию (ср. 281) нежелательна, увязка достижений современной экспериментальной акустической фонетики с фонологией кажется необходимой. Можно, конечно, не принимать во внимание эти результаты, а также теорию дифференциальных признаков фонем, как это делалось и до сих пор. Но в результате может оказаться только то, что знания русистов гарвардского университета в этой области будут более точными и более современными, чем знание наших. Вопрос только -- целесообразно-ли это?

Р.О. Якобсон по ходу всей своей плодотворной жизни был в центре или хотя бы в одном из центров (если не центром и одним из цетров ) революции, происходящей в языкознании в наши дни. Безусловно, что фонология была всегда передовой областью в этом процессе -- не в малой мере именно благодаря деятельности пражского лингвистического кружка и лично Р.О. Якобсону. Поэтому этот том невозможно не читать не только без интереса, но и без волнения. С таким же интересом ожидаем дальнейших томов избранных работ -- по морфологии, по истории филологии и по поэтике.

Ф. Пап

I Ср.: С.К. Шаумян, Преобразование информации в процессе познания и двухступенчатая теория структурной лингвистики. -- Проблемы структурной лингвистики. Отв. ред.: С.К. Шаумян. М. 1962. -- Стр. 5. -- Ср. также Предисловие к этому сборнику, написанное С.К. Шаумяном и его другие работы.

2 Ср.: "От иных семиотических систем естественные языки отличаются прежде всего семиотической двуплановостью: в естественных языках над первичной семиотической системой знаков надстраивается вторичная семиотическая система диакритических элементов знаков". -- С.К. Шаумян, Естественный язык как семиотическая система. -- Симпозиум по структурному изучению знаковых систем. Тезисы докладов. М. 1962. -- Стр. I3.

3 Не входим здесь в разбор вопроса о том, целесообразным является ли рассматривать буквы как знаки фонем. Р.О. Якобсон и в других местах указывает на то, что он считает и с точки зрения языка как абстрактной системы первичной звуковую сторону (см.: 474-5, 653 и т.д.). -- Наши, противоположные этому, взгляды, без ссылки однако на доводы Р.О. Якобсона см.: Slavica III Debrecen I963 , 2I sq.

4 См., напр., выступление одного из наиболее авторитетных руководителей упорядочения русской графики и орфографии: Д.Э. Розенталь, О реформе русского правописания. -- Русский язык в национальной школе. 1964, 2:3-I3, с ссылкой на выступление А.Б. Шапиро, М.В. Панова, но еще без ссылки на эту статью Р.О. Якобсона.

5 С точки зрения теории алгоритмов идеален был бы алфавит, в котором - между прочим - всякая буква была бы связной, т.е. "должна быть изобразимой без отрыва карандаша от бумаги" (см.: А.А. Марков, Теория алгорифмов. Москва-Ленинград, 1954 -- Стр. I3.) И хотя"алфавит"в теории алгоритмов или "слово" в этой же теории только слегка напоминают то значение, которое обычно в лингвистике придается этим

терминам, с точки зрения механической обработки информации, изложенной буквами какого-нибудь естественного языка, было бы неплохо, если бы мы могли быть ближе к идеальному алфавиту в смысле теории алгоритмов. А знак ь, так как он всегда указывает только на мягкость предшествующего звука, в данный момент считается элементом, который всегда "связно", т.е. вместе пишется с предшествующим письменным знаком. (Это действительно и в случаях написания в роде статья, вьюга.)

Первый програмированный учебник венгерского языка

/Szende Aladár: Magyar nyelvtan és fogalmazás a dolgozók iskolájának 8. osztálya számára. I. -- 115 p., II. -- 1o8 p. Budapest 1964./ А. Сенде, Венгерская грамматика и сочинение. Для 8-го класса вечерней школы.

Известно, что программированное обучение, начатое фактически еще во второй половине двадцатых годов /ср. первую статью одного из пионеров программированного обучения, С.Л. Пресси, вышедшую в 1926 г./, в последние годы все шире и шире распространяется в разных странах. Также известна связь этого дидактического метода как с формальным анализом языка /если речь идет о программированном обучении языку/, так и вообще с вопросами механизации умственных процессов. Программированный учебник можно рассматривать как бы в качестве аналога программы для машины. В первом дается система команд естественному автомату -- человеку, в виде своеобразного алгоритма; во втором -- это же самое дано автомату в более узком смысле этого слова /т.е. "искусственному" автомату/. Обучение человека и обучение машины в каждом случае имеет много общего между собой; в случае программированного обучения это сходство еще более наглядно. По всему поэтому мы с радостью приветствуем первый программированный учебник венгерского языка; этот учебник по нашим сведениям является одновременно и первым программированным учебником <u>вообще</u> в Венгрии.

Как видно из самого заглавия -- рецензируемый учебник предназначен для школ взрослых /для участников вечерней школы/. Учитывая условия, в которых происходит преподавание в подобных школах, можно считать очень удачной попытку применять этот метод впервые именно там. После положительного опыта в вечерней школе, если и с известными модификациями, очевидно, более успешно можно будет внедрить этот метод в обычные школы. Уже в самом начале отметим, что опыт первых месяцев применения данного учебника в вечерней школе был весьма положительным. Участники вечерней школы с удовольствием пользовались новым пособием, с интересом "играли" в постепенном разгадывании усваиваемого ими материала. В самом начале сам метод казался некоторым трудным /"не понимали, что надо делать"/, но впоследствии и они пошли быстрыми темпами вперед и результаты проведенных контрольных работ были очень хорошими.

С о д е р ж а н и е двух томов определяется програмой для 8-го класса вечерних школ: это, в основном, простое /распространенное и нераспространенное/ и сложное /сложно-сочиненное и сложно-подчиненное/ предложение. Кроме этого в начале первого тома есть некоторые непрограммированные страницы по повторению пройденного в низших классах материала, а в конце как первого, так и второго тома -- предписанные по програме сведения о правилах сочинения, о правилах оформления некоторых документов /как: автобиографии, приглашения и т.д./. В дальнейшем мы будем рассматривать только стержневую часть книги, полностью программированную. /Эта часть занимает около 90 % всего объема./

П о с т р о е н и е п р о г р а м м и р о в а н н ы х ч а с т е й следующее: Сначала идет программированный "параграф" -- целостное единство из усваиваемого материала. Такие главы: подлежащее, сказуемое, обстоятельства и т.д. После каждой главы, различной по длине в зависимости от содержания, следуют повторительные упражнения по пройденному в данной главе материалу. В конце второго тома, естественно, даются повторительные упражнения по всему пройденному материалу.

П о с т р о е н и е о т д е л ь н ы х г л а в характеризуется следующими чертами. Каждая страница разделена на два, неравные между собой столбца. В левом, большем столбце приведены правила, примеры, правила с пропусками, примеры с пропусками и т.д. После /довольно тонкой/ вертикальной линии, проходящей через всю страницу, в правом, меньшем столбце даны "ответы" -- то, чем надо примеры, правила дополнить, ответы на поставленные слева вопросы и т.д. Следовательно, на правой стороне в большинстве случаев расположены лишь отдельные слова.

В н у т р е н н я я с т р у к т у р а отдельных глав может быть различной. В некоторых случаях прямо с примеров /или же с примеров с пропусками/ самим учащимся предлагается составить правило, которое однако в таком случае всегда полностью повторяется на левой стороне; да подсказывается оно предварительно не только примерами, но и частью формулировки правила. Иными словами, пользуясь символикой Эванса-Хомма-Глейзера, после некоторых eg или eg может следовать ru, повторяемое в левой части как ru. [I] В других же случаях, более часто, работа начинается с сообщения правила полностью, потом следуют примеры с пропусками, потом новое правило, примеры на него с пропусками и т.д. В конце каждой главы дается свод пройденных там правил, в виде правил с пропусками.

Проиллюстрируется метод автора на показе чрезвычайно типичной венгерской грамматической конструкции -- притяжательной конструкции. Предварительно следует только заметить, что по традиционной венгерской грамматике притяжательные конструкции являются разновидностью определительных конструкций, в буквальном переводе они называются притяжательно-определительными сочетаниями. Проходятся определительные сочетания во втором томе, в конце простого предложения. Непосредственно до притяжательно-определительных конструкций были пройдены две другие разновидности определительных сочетаний: качественно-определительные и количественно-определительные. После же притяжательно-определительных следуют немедленно сведения уже о сложном предложении.

Рассматриваемая глава помещена на стр. 31-38 второго тома. Первые полторы страницы введены примерами с пропусками. По ходу решения этих примеров учащиеся активно знакомятся с основными понятиями, необходимыми для них в дальнейшем /обладатель, обладаемое, "как ставится вопрос на этот член предложения?" =чей? /. Это именуется частью А главы. Дальше, вплоть до стр. 37, следует часть Б: данные в готовом виде пять правил и последующие за каждым из них "примеры с пропусками". Наконец, на последней странице этой главы, в части В даны эти же правила подряд, но в несколько измененном порядке и каждое -- с пропуском. Следует отметить, что А, Б, В -- **условно данные нами обозначения,** горизонтально глава так не разделяется. Более того: местами имеются горизонтальные линии, такие же тонкие, как и главная вертикальная линия, но они слишком мелко подразделяют материал, даже внутри одного правила и примеров на него с пропусками -- несколько полей, отделенных друг от друга этими горизонтальными чертами. /Новое **правило же не всегда** расположено после одной из таких горизонтальных линий: в дальнейшем эти линии **не будут воспроизведены нами.**/

Отрывок из части А. /самое начало главы/:

"Если мы читаем словосочетание:

молоток монтера, a gépész kalapácsa

то из этого нам ясно, кому принадлежит орудие. молоток

ég /В данном случае.../. Молоток принадлежит ... монтеру

Иными словами: молоток является предметом,

ég обладаемым монтером, монтер же является ... обладателем

... молотка.

Итак, в словосочетании Молоток монтера

$\widetilde{eg}$ монтер является ......., а молоток является ....... — обладателем, обладаемым"

Отрывок из части Б. /первое правило и первые примеры на него с пропусками/:

"В подобных словосочетаниях слово, означающее
$ru^1$ обладателя, называется в предложении притяжательным определением. Следовательно, в словосочетании
$\widehat{eg}$ Щипцы монтера притяжательным определением является... — монтера
В словосочетании Щипцы слесаря ставится вопрос на притяжательное определение так: ......, — чьи /щипцы/
следовательно, притяжательным определением
$\widetilde{eg}$ является ...... — слесаря"

/Следует заметить, что оба приведенные здесь примеры с пропусками фигурировали уже и раньше, в части А./

И, наконец, дается только структура части В.:
$\tilde{ru}^1$, $\widehat{eg}$, eg, $\tilde{ru}^2$, $\widetilde{ru}^3$, $\tilde{ru}^4$, $\tilde{ru}^5$

Для полного понимания метода, примененного А. Сенде, надо добавить, что в венгерской грамматической терминологии "обладатель" и "притяжательное определение" образовано из одного корня / birtokos, birtokos jelző /, итак, последний термин звучит как "определение обладателя". Поэтому можно навести учащихся на сам термин /т.е. на метаязыковой знак/ путем подчеркнутого применения несколько раз слова из естественного языка. Естественно, что при этом нарочно и постоянно смешиваются не только метаязыковые и просто языковые явления, но и метаязыковые и логические термины /ср. "обладатель" -- логика, "определение обладателя" -- метаязык/. Последнее наше замечание уже пролило некоторый свет на то, ч е м у собственно учат програмированным методом в данном учебнике. Это -- сугубо традиционная грамматика. Однако, на это не собираемся здесь распространяться, ибо, в конце концов, программированным методом можно учить и закон божий, сам метод при этом не при чем; да и странно было бы обвинять автора в том, что ничуть от него в данном случае не зависит /ведь программы учебников составляются заранее, были они составлены не одним А. Сенде, они - т.е. программы -- утверждены министерством и не подлежат изменению и т.д./. Что выгодно в этом методе -- вопиющие внутренние противоречия и несостоятельность т.н. традиционной грамматики благодаря ему становятся более наглядными и ощутимыми как раз потому, что здесь дан в возможно оголенном виде весь ход рассуждений грамматиста.

В заключение повторяем сказанное в самом начале: всячески можно только приветствовать эту книгу, составленную с фантазией и умением. Надеемся, что это только первая, но далеко не последняя ласточка программированных учебников в нашей стране. Ведь безусловно, что программирование людей может давать полезный опыт программированию машин также. Немаловажным является и то обстоятельство, что программированный учебник, в роде рецензируемого, полностью программируем и в машину. т.е. с применением подобных учебников становится возможным применение обучающих машин также в Венгрии.

Ф. Пап

[I] ru = правило, eg = пример; ru = правило с пропуском, eg = пример с пропуском. А. Сенде нигде не пользуется этой символикой, но его текст легко разбивается на эти единства.

# Машинный перевод в народной республике Болгарии

А. Людсканов

Осенью 1964 г. при секции "Математическая статистика и теория вероятностей" Математического института с Вычислительным центром БАН была создана группа "Математическая лингвистика и машинный перевод".[I]

До конца 1964 г. обсуждались и уточнялись рабочие и перспективные планы группы и началась экспериментальная работа. Были намечены следующие основные задачи: составить словарь для МП математических текстов с русского языка на болгарский разработан словник примерно в 2500 основ на базе текстов из области теории чисел; предполагается постепенно расширить этот словник и сделать переводящую часть независимой от входной; разработаны параметры словарной информации и принята стандартная форма записи ; разработать общую форму алгоритма лексического анализа; составить схемы пословного перевода русских научно-технических текстов и конкретно выяснить вопрос о том, каковы максимальные возможности пословного и поморфемного перевода, обусловленные спецификой соотношения русского и болгарского языка; разработать схемы "Порядок слов" и "Членные формы" в качестве первых этапов создания независимого болгарского синтеза; провести опытный перевод общественного-политического текста разрабатывается словарь примерно в 1000 основ и анализирующий алгоритм на базе комбинированного морфологического и структурно-синтаксического подхода; в дальнейшем предполагается перестроить этот алгоритм на базе чистого структурно-синтаксического анализа, а затем на базе семантического анализа и установить конкрктные преимущества каждого из этих вариантов при русско-болгарском МП .

---

[I] Состав группы: руководитель к.ф.н. н.с. А.Людсканов и трое математиков

Все это создает предпосылки для установления оптимальной /с точки зрения болгарского языка/ системы зависимого анализа русского текста и перехода к разработке независимого болгарского синтеза и перевода на болгарский язык с других языков.

Параллельно с этим ведется экспериментальная работа на машине Минск 2. Проведен пробный перевод нескольких предложений русского научного текста /из книги "Галилей", Б.Г.Кузнецова, изд. АН СССР, 1964/ и Подгатавливается Перевод французского научного текста / из т. 2.

1958/. Экспериментированы на сравнительно широком материале следующие схемы раздела I алгоритма лексического анализа "Членение входного текста на предложения / /". К схеме разработан небольшой словарь сокращений. В результате работы схемы входный текст сегментируется по чисто формальным признакам на предложения /под предложением понимается тот символ или та линейная последовательность символов, которая сегментируется в качестве такового процедурой, называемой "схема сегментирования на и последовательно каждое предложение записывается в оперативной памяти, начиная с определенного адреса. Вторая схема "Членение рабочего предложения на слова " срабатывает после первой. Эта схема разработана в двух вариантах: первый вариант более простой - при нем при перфорации входного текста перед всеми знаками препинания, неотделенными от предшествующих или последующих пробелами, дается пробел; при втором варианте текст перфорируется без каких бы то ни было изменений. В результате работы этой схемы рабочее предложение сегментируется на слова /под словом понимается тот символ или та линейная последовательность символов, которая сегментируется в качестве такового процедурой, называемой "схема сегментирования на " / и каждое слово записывается, начиная с определенного адреса оперативной памяти, в четырех последовательных клетках. Обе схемы состоят из 106 приказов и занимают 132 клетки. Третья схема - это схема "Собственные имена". Схема срабатывает после поиска в словаре. Вводится понятие "ненайденный элемент". По формальным признакам машина устанавливает является ли этот элемент формулой, ненайденным словом или собственным именем. В первом случае дается приказ печатать, во втором в последовательности предложения оставляется цифровой эквивалент русского слова, и машина запоминает его для последующего пополнения словаря, а в третьем случае машина констатирует наличие собственного имени и начинает его обработку. Однобуквенные или многобуквенные сокращенные собственные имена печатаются на своем месте в последовательности предложения, а несокращенные подвергаются морфологическому анализу. Путем -операции, т.е. справа налево отсекается окончание, идентифицируется в морфологическом словаре, и его переводное соответствие прибавляетс к основе. Кроме этого в настоящее время в качестве дипломных работ разрабатываются и следующие схемы: схема "Чередованные основы существительных", схема "Устойчивые сочетания" и схема "Сложные слова" на базе русских математических текстов.

Проведенная экспериментальная работа создана предпосылки для постановки и решения двух принципиальных вопросов. Установить, исходя из рационального сочетания теоретических и утилитарных требований, следует ли, и если да, то в какой степени, допустить вмешательство того человека, который будет перфорировать входный текст /сегментация на предложения, сегментация на слова, отделение знаков препинания пробелами, известная обработка формул, иноязычных цитат и пр./. Второй вопрос сводится к созданию набора стандартных подпрограми для МП. Кроме этого в результате сравнения отдельных словарей предполагается выделить т.н. общую лексику и применить к ней принцип компримированной записи.

Параллельно с непосредственной работой по МП в настоящее время ведется и ряд языковых исследований, статистических исследований и работа в некоторых смежных областях.

В Секции славянских языков при Институте болгарского языка БАН ведется работа по установлению длины слов. Кроме этого руководитель секции проф. И.Леков исследовал проблемы валентности фонем, морфем и слов в славянских языках и разрабатывает вопросы связанные с сущностью и проявлениями синтаксической структуры. Предлагается понятие синтаксемы в качестве основной синтаксической единицы и обсуждается вопрос об изомеризме языковых ярусов.

Сотрудник института ст.н.с. Р.Мутафчиев разрабатывает проблемы статистического распределения частей речи в болгарских текстах в качестве стилистической характеристики.

продолжает свою деятельность созданный в 1962 г. лингвистический семинар при секции "Математическая статистика и теория вероятностей" МИ БАН /под руководством доц.Б.Пенкова и доц.Б.Сендова/, на котором читаются и обсуждаются научные доклады, представляющие интерес как для лингвистов, так и для математиков.

Кроме этого коллектив под руководством доц. М.Янакиева /доц. Б.Пенков, доц. Б.Сендов, ст.н.с. Р.Мутафчиев, к.ф.н. н.с. А.Людсканов и Д.Карабаджакова/ начал работу над составлением фреквентного словаря болгарского языка с использованием сортирующей машины Аритма. Подготовлена формализованная инструкция для сегментации входного текста на фонетические слова и нанесения некоторой дополнительной информации /означение фразового ударения и пауз, фонематическая транскрипция, учитывающая позиционные варианты фонем и пр./. Обособленные таким образом фонетические слова и дополнительная информация будут записаны на перфокартах. В результате работы сортирующей машины будет получен словник фреквентного словаря с указанием честот наиболее встречаемых фонетических слов и данные относительно статистического распределения фонетических слов различной длина в болгарских текстах.

Сотрудник Секции "Математическая статистика" Ем.Симеонов подготовил к печати работу об аналогии между лингвистической и термодинамической статистикой.

На кафедре болгарского языка филологического факультета СУ разрабатываются дипломные работы в связи с фреквентным словарем болгарского языка. К концу 1965 г. предпологается накопить информацию о честотах слов в тексте с общей длиной в 3 000 000 графических словоформ.

В Институте для Иностранных студентов /София/ делаются первые шаги к разработке проблем програмного обучения.

Начиная с 1965 г. проблемы МП включены и в исследовательские планы ЦИНТИ при Комитете по делам научного и технического прогресса.

В дальнейшем предпологается в значительно большей степени координировать и объединить усилия Группы МП, секции "Математическая статистика и теория вероятностей" ВЦ МИ БАН, Института болгарского языка БАН и в первую очередь Секции славянских языков, кафедр филологического факультета, Института для иностранных студентов и ЦИНТИ в деле расширения работы по МП и в смежных областях. Эта работа будет скромным вкладом в дальнейшее развитие нашего марксисткого языкознания.

Из.Леков, Към въпроса за съчетаемостта /валантността/ на частитна речта в българското изречение.

1962, стр.47-57.

Р.Мутафчиев, Някои статистически наблюдения върху стила на мемоара в последното десетилетие на миналия век, Славистически сборник, БАН, С., 1963; Статистическото разпределение на глагола като стилистическа характеристика на мемоара през последното десетилетие на миналия век, ЕЛ /в печати/.